

Tekoälyn etiikkaa

Valtiot tekoälyn hyödyntäjinä

Henrikki Pöntinen, 014341043
Teologisen etiikan ja sosiaalietiikan tutkielma
Joulukuu 2019

| | | |
|--|-----------------------------------|--|
| HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET | | |
| Tiedekunta/Osasto – Fakultet/Sektion Teologinen tiedekunta | | Laitos – Institution Systemaattinen teologia |
| Tekijä – Författare Henrikki Pöntinen | | |
| Työn nimi – Arbetets titel Tekoälyn etiikka - Valtiot tekoälyn hyödyntäjinä | | |
| Oppiaine – Läroämne Teologinen etiikka ja sosiaalietiikka | | |
| Työn laji – Arbetets art Maisterin tutkielma | Aika – Datum 21.12.2019 | Sivumäärä – Sidoantal 69 |
| <p>Tekoälyteknologia mahdollistaa paljon asioita. Sen avulla voidaan muun muassa kehittää uusia ratkaisuja sosiaaliin ja ekologisiin ongelmiin, tehostaa teollisuuden alojen toimintaa, avustaa päätöksentekoprosessien perustamista empiiriseen informaatioon ja tuottaa ihmisten yksilölliset piirteet huomioon ottavia palveluita. Kehitykseen liittyy kuitenkin myös riskejä. Tekoälyavusteinen päätöksenteko voi sisältää epätasa-arvoa voimistavia vinoumia, tekoälyteknologiaan käytettävä herkkäluonteinen data voi altistua väärinkäytöksille ja tekoälyteknologioita voidaan käyttää epäilyttäviin tarkoitukseen, kuten totalitaaristen valtioiden kontrollin lisäämiseen. Tekoälykehitys on hajautunut ja sen mahdollistamia uusia innovaatioita on vaikea ennustaa.</p> <p>Tarkastelen tutkielmassani valtioita tekoälyn hyödyntäjinä etiikan näkökulmasta. Vastaan kysymyksiin kuten, mitä tarkoittaa, että valtiot hyödyntävät tekoälyä ja millaisia velvollisuuksia ja vastuita siihen sisältyy. Tekoälyn valtiollisen hyödyntämisen tarkastelun ajankohtaisuutta korostaa se, että vuoden 2017 ja 2019 välillä 25 valtion hallitukset ovat julkaisseet kansallisen tekoälystrategian. Analysoin tutkielmassani Ison-Britannian, Ranskan, Suomen ja Yhdysvaltojen hallitusten tekoälystrategioita kehittäneiden työryhmien raportteja, ja vertaan niiden heijastamaa eettistä pohdintaa tekoälyn yhteiskunnallisesta hyödyntämisestä käytyihin tieteellisiin keskusteluihin. Hyödynnän raporttien vertailussa systemaattista analyysiä, mikä tarkoittaa käsite-, argumentaatio- ja edellytysten analyysin yhdistelmää.</p> <p>Tekoälyn etiikka voidaan jakaa käsiteanalyysin avulla kolmeen alakategoriaan, jotka ilmentävät tekoälyn etiikan sisältämiä vastuusuhteita. Kategoriat ovat ohjelmoinnin etiikka, dataetiikka ja hyödyntämisen etiikka. Ohjelmoinnin etiikka ja dataetiikka ovat alisteisia hyödyntämisen etiikalle, joka vastaa kysymyksiin kuten, mitä arvoja tekoälyn hyödyntämisellä halutaan tavoitella ja miten hyödyntäminen vaikuttaa yksilöön, yhteisöihin ja laajemmin yhteiskuntaan.</p> <p>Maija-Riitta Ollilan, Erkki Laitilan ja Yuval Hararin mukaan inhimillinen tekoälyn kehitys ja hyödyntäminen edellyttävät avointa ja laajaa arvokeskustelua. Ilman sitä tekoälyn kehitys tapahtuu piilossa useimmilta ihmisiltä ja markkinatalouden mekanismien muokkaamana. Kyseiset mekanismit ohjaavat kehitystä lyhyen tähtäimen tavoitteiden mukaisesti, mikä on tekoälyn eettisen hyödyntämisen periaatteiden vastaista. Ranskan ja Suomen hallitusten työryhmien raportit antavat viitteitä siitä, että tekoälyn kehitystä voidaan ohjata itseisarvojen, kuten oikeudenmukaisuuden kautta, vaikka yksi kehityksen tavoitteista olisi taloudellisen hyödyn saavuttaminen.</p> <p>Hyödyntämisen etiikan toteutumisen edellytys on, että valtion hallinto tunnistaa erityisasemansa tekoälyn vuorovaikutteista kehitystä ohjaavana tahona. Valtion hallinto kykenee arvokeskusteluiden fasilitoinnin ja tukemisen, lainsäädännön, koulutuksen tarjonnan, kansainvälisten yhteistyösopimusten, rahoituksen kohdentamisen ja julkisten hankintojen ja palveluiden tarjonnan kautta vaikuttamaan laaja-alaisesti yhteiskunnallisen ekosysteemin toimintaan. Erityisasiemiansa vuoksi on tärkeää, että valtion hallinnot eivät jää passiivisiksi toimijoiksi tekoälyn etiikan kehittämisen osalta.</p> | | |
| Avainsanat – Nyckelord Tekoäly, Tekoälyn etiikka, Yhteiskuntaetiikka, Creating Shared Value | | |
| Säilytyspaikka – Förvaringställe Helsingin yliopiston kirjasto, Keskustakampanuksen kirjasto, Teologia | | |
| Muuta tietoa | | |

Sisällysluettelo

| | |
|---|----|
| 1 Johdanto | 1 |
| 1.1 Tekoälyn määritelmä..... | 4 |
| 1.2 Tekoälyn etiikka | 7 |
| 1.3 Vastuun tasot..... | 12 |
| 2 Arvot tekoälyn hyödyntämisen ohjaajina..... | 16 |
| 2.1 Kehitystä kehityksen vuoksi | 17 |
| 2.2 Systeemievoluutio | 22 |
| 2.4 Ihmisen ja tekoälyn yhteys | 25 |
| 2.5 Markkinoiden voimalla päin seinää..... | 27 |
| 3 Antroposeenin etiikka | 31 |
| 3.1 Yhteistä toimintaa rakentamassa | 33 |
| 3.2 Aikaraamin vaikutus arvoihin..... | 38 |
| 4 Strategioiden ilmentämä etiikka | 40 |
| 4.1 Kansalaisten hyväksyntä minimiarvona..... | 41 |
| 4.2 Eettisyys kestäväenä pohjana..... | 46 |
| 4.3 Strategioiden ekosysteemit | 48 |
| 4.4 Eettisyys kehityksen hidasteena | 52 |
| 5 Loppukatsaus | 57 |
| Lähde- ja kirjallisuusluettelo..... | 61 |
| Käytetyt lyhenteet | 61 |
| Lähteet ja apuneuvot | 61 |
| Kirjallisuus | 61 |
| LIITTEET | 70 |
| Liite 1. Tekoälystrategiat taulukoituna | 70 |
| Liite 2. Strategioiden tiivistelmät..... | 71 |

1 Johdanto

Tarkastelen tutkielmassani valtioita tekoälyn hyödyntäjinä eettisestä näkökulmasta. Mitä tarkoittaa, että valtiot hyödyntävät tekoälyä? Millaisia velvollisuuksia ja vastuita siihen sisältyy? Tekoälyn yhteiskunnallisista vaikutuksista ja eettisestä hyödyntämisestä on tähän mennessä tehty melko vähän tieteellistä tutkimusta. Etenkin Euroopassa on viime vuosina havahduttu aiheen tärkeyteen. EU:n komission Horisontti 2020 tutkimus ja innovaatio -hankkeen yksi teema on ollut levittää tietoisuutta *Responsible Research and Innovation* (RRI) -viitekehyksestä. RRI:n tavoite on saada tutkijat ja insinöörit sisällyttämään yhteisöjen odotuksia ja kehitettävien teknologioiden sosiaalisten vaikutusten arviointia teknologioiden suunnitteluvaiheeseen. Tällä tavoin teknologiat kehitettäisiin suunnittelusta lähtien yhteisöjä varten ja sosiaalisesti kestäväällä tavalla. Suomessa Valtionneuvoston kanslia on rahoittanut RRI:n mukaista tutkimusta tekoälyn eettisestä hyödyntämisestä viranomaistoiminnassa. Lisäksi Valtiovarainministeriö on perustanut *Aurora* -tekoälyohjelman, jonka tavoite on valmistella julkisten palveluiden ihmislähtöistä siirtymistä tekoälyaikaan.¹

Vaikka tekoälyn yhteiskunnallisten vaikutusten tutkimusta etiikan näkökulmasta on tehty vasta vähän, aiheen tärkeys on tunnustettu. Tekoälyn mahdollisuuksista ja riskeistä julkaistaan lähes päivittäin uutisartikkeleita niin kansallisissa kuin kansainvälisissäkin tiedotusvälineissä. Työtehtävien automatisoinnin kiihtyminen, herkkäluonteisen datan väärinkäyttö ja algoritmivinoumat² sekä tekoälyn mahdollistamat uudet teknologiat ovat keskusteluiden keskiössä. Vuoden 2019 aikana on huolestuttu kansalaisten valvomisen ja vainoamisen mahdollisuuksista, joita tekoäly tarjoaa totalitaarisille valtioille.³ Tekoälykehitys on hajautunut ja sen mahdollistamia uusia innovaatioita on vaikea ennustaa. RRI osoittaa, että valtiollisilla toimijoilla on mahdollisuus vaikuttaa tapaan, jolla tekoälyä kehitetään ja hyödynnetään, jos he tunnistaisivat tarpeen pohjustaa ja tukea yhteiskunnallista keskustelua kehityksen toivottavasta suunnasta.⁴

Tekoälyn valtiollisen hyödyntämisen tarkastelun ajankohtaisuutta korostaa se, että vuoden 2017 ja 2019 välillä 25 valtion hallitukset ovat julkaisseet kansallisen

¹ Koivisto et al. 2019; Kansallisen AuroraAI-tekoälyohjelman toimeenpano s.a.; PERRI s.a..

² Algoritmivinoumalla kuvataan ilmiöitä, joissa tekoälyn päätöksenteko ilmentää ihmisten ennakkoluuloja tai päättelyvirheitä. Vinoumat aiheutuvat, kun tekoälyn kehittämisen ja päätöksenteon pohjana käytetään dataa, joka sisältää ihmisten ennakkoluuloja tai kontekstiin sopimatonta materiaalia. Myös algoritmin kehittäjät siirtävät tahattomasti omaa ajatusmaailmaa osaksi algoritmeja, mikä mahdollistaa vinoumien muodostumisen. Hern 2018; Ollila 2019, 120.

³ Ahlroth 2018; Asikainen 2018; Heikkinen 2018; Hern 2018; Koutonen 2018.

⁴ Laitila 2019, 19.

tekoälystrategian.⁵ Analysoin tutkielmassani Ison-Britannian, Ranskan, Suomen ja Yhdysvaltojen hallitusten tekoälystrategioita kehittäneiden työryhmien raportteja, ja vertaan niiden heijastamaa eettistä pohdintaa tekoälyn yhteiskunnallisesta hyödyntämisestä käytyihin tieteellisiin keskusteluihin. Valitsin vertailtavat valtiot sen perusteella, että niillä on todellinen mahdollisuus vaikuttaa tekoälyn kehitykseen ja hyödyntämiseen yksittäisinä kansainvälisinä poliittisinä toimijoina tai yhteistyössä muiden kanssa.⁶ Valintaan ovat lisäksi vaikuttaneet kyseisten valtioiden maantieteelliset ja poliittiset erot, jotka vaikuttavat muun muassa siihen, onko valtio osa jotain suurempaa yhteistyökuviota, kuten esimerkiksi EU:n tai Baltian ja Pohjoismaiden tekoälyohjelmia. Erot tekevät vertailujoukosta heterogeenisemmän tarkastelukohteen.⁷

Hyödynnän raporttien vertailussa systemaattista analyysii, mikä tarkoittaa käsite-, argumentaatio- ja edellytysten analyysin yhdistelmää. Argumentaatioanalyysin ja edellytysten analyysin avulla redusoin vertailtavat raportit argumenttien tasolle ja konstruoin ne uudelleen tutkimuskirjallisuudesta jäsentämäni viitekehysten avulla.⁸ Metodini tarkoitus on tehdä raporttien sisältämät taustaoletukset näkyviksi, jotta niiden oikeutettavuutta ja todenmukaisuutta voidaan arvioida. Tutkielmani tuottaa lisätietoa siitä, mitä eettinen hyödyntäminen tarkoittaa tekoälykehityksen poliittisessa hallinnoinnissa.

Jotta voin puhua tekoälyn etiikasta, täytyy minun ensiksi määritellä, mitä tekoälyn käsite tarkoittaa. Määrittelyn tarpeellisuutta korostaa se, että termille ei ole vakiintunut yhtä kuvausta edes akateemisissa keskusteluissa, vaikka yhteneviä linjoja on hahmotettu. Paitsi tekoälyn myös tekoälyn etiikan käsitettä käytetään kuvaamaan laajaa asioiden kirjoa.⁹ Jäsennän määrittelyä varten aiheista käytyä tieteellistä keskustelua. Lopuksi hyödynnän käsiteanalyysii hahmottaakseni käsitteiden keskeisiä tekijöitä ja niiden välisiä suhteita sekä toisiinsa että tekoälyn ymmärtämiseen. Määritelmien avulla on helpompi tarkastella, mitä tekoälyn hyödyntämisellä tarkoitetaan, mutta ne auttavat myös tutkimuskirjallisuuden ja aineistona käyttämäni raporttien vertailussa ja analysoinnissa. En syvenny tutkielmassani yksittäisten tekoälyteknologioiden tarkasteluun, vaan käsittelen aihetta tekoälyn kattokäsitteen kautta.

⁵ Dutton 2018.

⁶ Katso muun muassa Bughin et al. 2017, 15.

⁷ Alkuperäinen tarkoitukseni oli sisällyttää useamman mantereiden valtioiden strategioita vertailuun, mutta tutkielmani rajallisen pituuden vuoksi päätin rajata vertailujoukon koskemaan länsimaisia valtioita.

⁸ Hallamaa s.a.; Jolkkonen 2007, 12.

⁹ Ollila 2019, 8–11.

Tutkielmani ensimmäisessä osiossa keskityn tekoälyn ja tekoälyn etiikan käsitteiden määrittelyyn. Toinen ja kolmas osio koostuvat kirjallisuuskatsauksista, joissa laadin tekoälyn eettisen hyödyntämisen viitekehyksen käyttämällä muun muassa Maija-Riitta Ollilan, Erkki Laitilan ja Yuval Noah Hararin teoksia, sekä Creating Shared Value -teoriaa. Ollilan teos *Tekoälyn etiikkaa* (2019) keskittyy eettiseen pohdintaan, Hararin teos *Homo Deus, huomisen lyhyt historia* (2015) käsittelee tekoälyn kehitystä historian tutkimuksen näkökulmasta ja Laitilan teos *Ihanteeksi vastuullinen tekoäly* (2019) pohjautuu systeemitieteisiin. Creating Shared value on yhteiskuntavastuuta koskeva teoria, joka pyrkii kapitalistisen talousjärjestelmän systeemitasoiseen muuttamiseen. Kirjoittajien teoksia ja Creating Shared Value -teoriaa yhdistävät yhteiskunnallinen lähestymistapa aiheeseen ja ajatus, jonka mukaan tekoälyn ja tekoälyteknologioiden kehityssuunta riippuvat nyt käynnissä olevasta arvokeskustelusta.

Käytän laatimaani tekoälyn eettisen hyödyntämisen viitekehystä tutkielman neljännessä luvussa vertailemieni raporttien analysoinnissa. Tiivistän havaintoni tutkielman viidennessä luvussa.

1.1 Tekoälyn määritelmä

Tekoälyä tutkinut valtiotieteiden tohtori Maija-Riitta Ollila toteaa teoksessaan *Tekoälyn etiikkaa* (2019), että tekoälyn käsitettä ei voida määritellä tavalla, joka olisi opillisesti puhdas.¹⁰ Tämän vuoksi hän päätyy käsittelemään tekoälyä sen mukaan, mikä on yleisessä yhteiskunnallisessa keskustelussa hyväksytty tekoälyksi.¹¹ Pyrin itse määrittelemään tekoälyn tarkemmin, koska yhteiskunnalliseen keskusteluun perustuva määritelmä ei tarjoa mielestäni perustaa rakentaa systemaattista teoriaa. Epämääräinen määrittely voi johtaa muun muassa siihen, että käsitellään enemmänkin digitalisoitumista kuin tekoälyä, joka on osa digitalisaatioon vaikuttavista teknologioista. Käsitteiden määritelmät ovat yhteydessä niiden kuvaamien asioiden tuottamiin kokemuksiin. Määritelmät vaikuttavat siihen, mitä asioiden kannalta pidetään tärkeinä. Käsitteiden määrittely on yksi filosofisen tutkimuksen tärkeimpiä tehtäviä.¹²

Tekoälyn käsitteellä (engl. *artificial intelligence*) tarkoitetaan niin itse teknologisia ratkaisuja kuin niiden tutkimustakin. Maailmanlaajuisessa käytössä olevien tietojenkäsittelytieteiden oppikirjojen kirjoittajat, Peter Norvig ja Stuart Russell sanovat, että tekoälytutkimus pyrkii kognitiotieteiden tapaan ymmärtämään inhimillistä yleisälyä. Tekoälytutkimus on alkanut Norvigin ja Russellin mukaan jo vuonna 1943, vaikka tieteenalan nimi vakiinnutettiin vasta 1956 kansainvälisessä *Paramouthin* konferenssissa. Nimitykseen tekoäly päädyttiin sen vuoksi, että konferenssissa saavutetun konsensuksen mukaan tietokoneen avulla voitaisiin kehittää älykkäältä vaikuttava keinotekoinen toimija. Norvig ja Russell määrittelevät tekoälyn ”parhaan keinotekoisin agentin kehittämiseksi annettuun kontekstiin”.¹³ Määritelmässä tekoälyn kontekstisidonnaisuus on ilmeinen asia. *Paras* keinotekoinen lakiavustaja ei todennäköisesti ole paras itseohjautuvan auton kuljettaja. *Keinotekoinen* tarkoittaa jotain ihmisen kehittämää ja *agentti* toimijan kaltaisuutta. Toimijalla on perinteisesti kuvattu filosofiassa ihmisen kaltaista ajattelevaa olentoa. Norvigin ja Russellin määritelmän mukaan tekoälyllä on autonomisen ongelmanratkaisijan piirteitä.

Valtion teknologian tutkimuskeskuksen (VTT) *Tekoälyn käsitekartassa* (2018) esitetyssä tekoälyn määritelmässä on hyödynnetty Norvigin ja Russellin ajatuksia. Käsitekartan kuvauksen mukaan ”tekoälyn avulla koneet, laitteet, ohjelmat, järjestelmät ja

¹⁰ Ollilan teos pyrkii herättämään laaja-alaisesti ajatuksia tekoälyn etiikasta. Toisin sanoen popularisoimaan aihetta. Liian tarkka keskittyminen käsitelmääritelmään voisi vaikuttaa negatiivisesti popularisoinnin onnistumiseen.

¹¹ Ollila 2019, 9.

¹² Hallamaa 2017, 11.

palvelut voivat toimia tehtävän ja tilanteen mukaisesti järkevällä tavalla”.¹⁴ Määritelmän ongelma on, että se ulkoistaa tekoälyn joksikin koneiden, laitteiden, ohjelmien, järjestelmien ja palveluiden ulkopuoliseksi toimijaksi. Määritelmä saa kysymään, mistä tekoäly on tullut ohjaamaan laitteita ja koneita sekä mitä järkevällä tavalla toimimisella tarkoitetaan.

Työ- ja elinkeinoministeriön (TEM) julkaisussa *Suomen tekoälyaika – Suomi tekoälyn soveltamisen kärkimaaksi: Tavoite ja toimenpidesuosituks* (2017) –asiakirjassa esitetyssä määritelmässä ei ole samanlaista puutetta kuin VTT:n julkaisussa. Siinä tekoälyn kuvataan tarkoittavan laitteita, ohjelmistoja ja järjestelmiä, jotka kykenevät oppimaan ja tekemään päätöksiä lähes samalla tavalla kuin ihmiset.¹⁵ Tekoäly määritellään näin laitteiden, ohjelmistojen ja järjestelmien ominaisuudeksi. VTT:n ja TEM:n julkaisujen määritelmiä kuitenkin yhdistää se, että niissä käytetään kiertoilmausta älykkyyden kuvaamiseksi. Vaikka kaikki edellä mainitut määritelmät yhdistettäisiin täydentämään toisiaan, ei olisi vielä vastattu kysymykseen, miten elottomina pidetyillä asioilla voi olla nisäkkäiden ominaisuudeksi rajattu ominaisuus, älykkyys.¹⁶

Oxfordin yliopiston soveltavan etiikan professori Nick Bostrom ja *ystävällisen tekoälyn* teorian kehittäjä Eliezer Yudkowsky eivät artikkelissaan *The Ethics of Artificial Intelligence* (2011) määrittele suoraan, mitä he tarkoittavat tekoälyllä. Sen pohjalta voidaan kuitenkin konstruoida seuraava määritelmä: tekoäly ilmenee, kun algoritmi mahdollistaa dataa hyödyntämällä laitteen tai muun vastaavan ei-ihmisen älykkään toiminnan.¹⁷ Käytän tässä tutkielmassa Bostromin ja Yudkowskyn artikkelin pohjalta muotoilemaani tekoälyn määritelmää, jossa hyödynnän VTT:n käsitekartassa käytettyä tapaa kiertää älykkyyden määrittely. Tekoäly tarkoittaa tässä tutkimuksessa algoritmin ja sen hyödyntämän datan yhteisvaikutusta, joka mahdollistaa laitteen tai muun vastaavan ei-elollisen toimia jonkin tehtävän ja tilanteen mukaisesti järkevällä tavalla. Kyseisen määritelmän mukaan tekoäly on kokonaisuuden summa eikä sitä voida redusoida yksittäisiin osatekijöihinsä.

Määritelmä antaa hyvät lähtökohdat pohtia tekoälyyn liittyviä eettisiä kysymyksiä, sillä sen avulla tekoäly voidaan jakaa osatekijöiksi. Sen mukaan tekoälyn minimi-edellytyksiä ovat algoritmi(t), toiminta-alusta (tietokone, robotti tai muu vastaava laite) ja hyödynnettävä data. Määritelmä ei myöskään ole ahdas, minkä ansiosta sitä voidaan käyttää tekoälyn erilaisten ilmenemismuotojen tarkastelemiseen. Algoritmi on keskeinen osa hyödyntämääni

¹³ Norvig & Russell 2010, 1020.

¹⁴ Ailisto 2018, 4.

¹⁵ Tekoälyohjelman ohjausryhmä 2017, 62.

¹⁶ Harari 2017, 91; 133.

määritelmää. Sen vuoksi täsmennän myös algoritmin käsitteen. Norvig ja Russell kuvaavat, kuinka algoritmi on järjestelmällinen matemaattinen toimintaohjeiden sarja. Se antaa *keinotekoisille agenteille* ohjeet, mitä tehdään ja missä järjestyksessä. Yuval Noah Harari laajentaa Norvigin ja Russellin määritelmää teoksessaan *Homo Deus* (2015) sanomalla, että algoritmi on synonyymi toimintaohjeelle. Tämän määritelmän mukaan keitto-ohjeetkin ovat algoritmeja, vaikka niitä ei olisi kirjoitettu matemaattisen kaavan muotoon.¹⁸

Bostrom ja Yudkowsky puhuvat artikkelissaan tekoälyn eri kehitysasteista.

Yksipuoliseen älykkääseen tehtävään kykenevä tekoäly edustaa kapeaa tekoälyä (nykyinen kehitysaste), kun taas keinotekoinen yleisäly (*artificial general intelligence*), jonka älykkyys vastaa ihmisen laaja-alaista älykkyyttä, edustaa vahvaa tekoälyä. Kolmanneksi kehitysasteeksi he määrittelevät supertekoälyn (*artificial superintelligence*), joka ylittää ihmisten älykkyuden kaikissa merkityksellisissä yhteyksissä. Myös Norvig ja Russel käsittelevät heikon ja vahvan tekoälyn termejä kirjassaan *Artificial intelligence: modern approach 3rd edition* (2018). Heidän määritelmänsä on yksiselitteisempi kuin Bostromin ja Yudkowskyn, sillä he toteavat, että heikko tekoäly vain näyttää kykenevän ajattelemaan, kun taas vahva tekoäly todella ajattelee.¹⁹ Bostromin ja Yudkowskyn muotoilemat tekoälyn kehitysvaiheiden määritelmät ovat tarkempia ja selvempiä²⁰ kuin Norvigin ja Russellin. Hyödynnän tarvittaessa Bostromin ja Yudkowskyn määritelmiä.

Kehitysasteiden lisäksi tekoälyteknologioita voidaan jaotella sen mukaan, minkälaista ohjelmointitapaa niissä hyödynnetään. Norvig ja Russel mainitsevat kolme päälinjaa, jotka ovat logiikkaan, todennäköisyyslaskentaan ja keinotekoiseen neuroverkostoon pohjautuvat tekoälyt. Ne erottaa toisistaan se, että logiikkaan pohjautuva tekoäly edellyttää, että kaikki toimintaohjeet kirjoitetaan auki, mutta todennäköisyyksiin ja neuroverkostoon pohjautuvat tekoälyt eivät edellytä sitä. Tästä voidaan päätellä, että todennäköisyyslaskentaan (koneoppiminen) ja keinotekoiisiin neuroverkostoihin (syväoppiminen) perustuvat tekoälyt ovat reilusti logiikkaan (koneoppiminen) perustuvia tekoälyjä tehokkaampia, sillä ne voivat sivuuttaa osan vaiheista, jotka logiikkaan perustuvan tekoälyn täytyy käydä läpi.

¹⁷ Bostrom & Yudkowsky 2014, 1–5.

¹⁸ Hararin voi tulkita määrittelevän ihmisetkin perusolemukseltaan algoritmeiksi. Se on kuitenkin Hararin käyttämä tehokeino, jolla hän pyrkii kiinnittämään huomiota transhumanismin kaltaista teknologista kehitystä kannattavien henkilöiden taustaoletuksiin. Palaan Hararin määritelmän tarkoitukseen tarkemmin myöhemmissä luvuissa. Harari 2017, 91; Ollila 2019, 87.

¹⁹ Norvig ja Russell 2018, 1020.

²⁰ Norvigin ja Russellin määritelmän heikkous on, että se ei erota kapeaan toimintaan ja laajaan toimintaan kykeneviä vaiheita toisistaan eikä se anna työkaluja ajattelevan ja ei ajattelevan tekoälyn erotteluun.

Syväoppimisella pyritään luomaan ihmismäistä *ajattelua*.²¹ Eri ohjelmointitavat ja kehitysasteet ovat antaneet aiheen niille tyypillisille kysymyksille myös etiikan näkökulmasta, mutta tutkielmassani keskityn tarkastelemaan aihetta tekoälyn kattokäsitteen kautta.

Tekoälystä käytetään satunnaisesti myös sanaa keinoäly.²² Englanninkielinen termi *artificial* käännetään suomeksi sanoilla keinotekoinen, teko- tai keino-.²³ Vaikka termistö saattaa aiheuttaa hämmennystä, keinoäly ja tekoäly ovat toistensa synonyymeja. Kuten Russellin ja Norvigin määritelmästä käy ilmi, englannin kielessä artificial intelligence voi myös tarkoittaa tekoälyn tutkimusta. Tutkielmassani tekoäly tarkoittaa kuitenkin ensisijaisesti itse teknologiaa eikä sen tutkimusta.

1.2 Tekoälyn etiikka

Etiikka on tutkimusala, jossa pyritään ymmärtämään hyvän ja pahan käsitteitä ja niihin liittyviä kysymyksiä, kuten hyvän elämän ja oikeudenmukaisen toiminnan periaatteita ja toteuttamisedellytyksiä.²⁴ Moraalifilosofien keskuudessa ei ole yli kahden vuosituhannen aikana päästy yhteisymmärrykseen siitä, mitä hyvä ja paha ovat ja miten voidaan elää hyvä elämä. Esimerkiksi antiikin aikaan hyvä rinnastettiin usein onnellisuuteen, mutta ei päästy yhteisymmärrykseen siitä, kuinka onnellisuus saavutettaisiin. Muun muassa Platon ajatteli puhtaan tiedon saavuttamisen johtavan onnellisuuteen, kun taas Aristoteles näki hyveiden eli oikeanlaisten luonteenpiirteiden kehittämisen tukevan onnellisuuden saavuttamista.²⁵

Ensimmäisen kristillisen vuosituhannen ja keskiajan aikana käsitys hyvästä muuttui onnellisuutta abstraktimmaksi ja se kuvattiin usein Jumalan avulla. Näin pyrittiin vastaamaan ongelmaan, että hyvää ei voida määritellä tyhjentävästi. Etiikassa alettiin tarkastella hyvää sen ilmenemismuotoja havainnoimalla ja systematisoimalla. Esimerkkeinä hyvän ilmenemisestä pidettiin muun muassa ihmisen kokemaa nautintoa, rakkautta ja onnellisuutta. Tällä tavoin etiikan tutkimuksessa luovuttiin pyrkimyksestä määritellä hyvyys tyhjentävästi ja sen sijalle tuli yritys kuvailla sen ilmenemismuotoja ja niihin vaikuttavia tekijöitä.²⁶

Valistusajattelijat pitivät yksilön arvoa ja oikeuksia moraalifilosofian tärkeimpinä aiheina, minkä vuoksi valta ja sen väärinkäyttö sekä hallitsijan (valtion edustajana) velvollisuudet muuttuivat eettisen pohdinnan keskeisiksi kysymyksiksi. Kansalaisoikeuksien käsitteen kehittämisen myötä myös hyvän kansalaisen idea ja kansalaisuuden tekijöiden

²¹ Norvig ja Russell 2018, 1024.

²² Katso muun muassa Rossi 2017.

²³ MOT s.a..

²⁴ Oksanen et al. 2010, 7; Hallamaa 2017, 8–9.

²⁵ Aristoteles 2005; Platon 2007.

tarkastelu määriteltiin moraalifilosofian kannalta tärkeiksi aiheiksi. Empirismin vaikutuksesta moraalifilosofia ja uskonnollinen ajattelu erkaantuivat toisistaan, mikä vaikutti siihen, että moraaliteorioita alettiin kehittää sekulaarein käsittein. Etiikan tutkimuksessa hyvästä toiminnasta²⁷ ja sen ymmärtämisestä tuli entistä keskeisempi tutkimuskohde.²⁸

1900-luvun alkupuolella syntyi pragmaattisen filosofian koulukunta, joka korosti käytännön elämän roolia tiedon ja teorioiden muodostumisessa ja kehittämisessä. Pragmaattinen lähestymistapa omaksuttiin myös moraalifilosofiassa. Ihmisten toiminnan päämäärien tulkittiin ilmentävän heidän arvojaan ja vastavuoroisesti arvojen ohjaavan toimintaa.²⁹ Samalla vuosisadalla tapahtuneet maailmansotien kauheudet saivat ihmisten luottamuksen hyvyteen ja rationaalisuuteen horjumaan. Sen vuoksi analyyttisen etiikan arvo tutkimusaiheena heikkeni. Etiikan tutkimus sai kuitenkin uutta nostetta 80-luvulla, kun käytännöllisen etiikan arvostus kasvoi. Pragmaattisen filosofian tavoin käytännöllinen etiikka on luonteeltaan kontekstisidonnaista.³⁰

Analyyttisen etiikan oivallukset eivät ole hyödyttömiä käytännöllisen etiikan tutkimukselle, vaan niitä voidaan käyttää tunnistamaan moraalifilosofian keskeisiin käsitteisiin, kuten oikeudenmukaisuuteen, hyvään ja pahaan, oikeaan ja väärään sekä moraalisen toimintaan liittyviä kysymyksiä. Yksi tapa selvittää keskeiset kysymykset on tarkastella, mitkä aiheet ovat säilyttäneet merkityksensä läpi etiikan historian. Niitä ovat muun muassa hyvän ilmentymät, eli arvot, toimijuus, vastuu ja valta. Arvot kuvastavat toiminnan päämääriä, toimija toteuttaa toiminnan, ja valta sekä valtarakenteet ovat keskeisiä pohdittaessa oikeudenmukaisuutta.³¹

Vastuun käsite on tärkeä erityisesti normatiivisen etiikan teorioille. Normatiiviset teorit rakentuvat ajatukselle, että ihmiset ovat vastuullisia teoistaan, koska heillä on vapaa tahto³². Vastuu sisältää puolestaan ajatuksen velvollisuudesta, eli velvoitteesta tehdä tai jättää tekemättä jotain. Muun muassa velvollisuusetiikassa vapaa tahto asettaa ihmiset moraalisesti velvolliseksi pohtimaan toimintaansa ja sen seurauksia suhteessa muihin moraalisiin

²⁶ Katso muun muassa Augustinus 2003.

²⁷ Moraali sana polveutuu latinan kielisestä sanasta *mos* (monikossa *mores*), joka tarkoittaa tapoja, käyttäytymismalleja. Glare 1992.

²⁸ Russou 1933; Locke 1997; Hume 1998.

²⁹ Philström 2007.

³⁰ Airaksinen 1987, 17; Vainio 2016.

³¹ Airaksinen 1987, 21; Hallamaa 2017, 20; Ollila 2019, 81.

³² Vapaan tahdon filosofia on laaja filosofisen tutkimuksen alue. Nykyisin ihmisen vapaan tahdon ajatellaan olevan vähintään rajallinen, sillä ihmisen toimintaa ohjaavat biologiset, psyykkiset ja ympäristön vaikutukset. Wheeler 2018.

olentoihin.³³ Seurausetiikka painottaa velvollisuusetiikkaa yksioikoisemmin, että toimija on vastuussa siitä, että hänen tekonsa seuraukset tuottavat mahdollisimman paljon hyvää.³⁴ Maija-Riitta Ollila mukailee velvollisuusetiikkaa sanomalla, että ihmisen reflektiokyky asettaa ihmisille velvollisuuden arvioida toimintaansa. Kykenemme tarkastelemaan toimintaamme ”toisasteisesti”³⁵ eli pohtimaan oman toimintamme oikeellisuutta. Reflektoinnin tekemättä jättäminen olisi Ollilan mukaan välinpitämättömyyttä. Vaikka eri teorioista ei voi muotoilla yksimielistä vastausta vastuun muodoista ja toimijan velvollisuuksista, niistä ilmenee ajatus, että toimija on vastuussa teoistaan.³⁶

Analyttisen pohdinnan lisäksi etiikan historia tarjoaa metodeja, joiden avulla käytännön ongelmia voidaan tutkia. Deskriptiivisen etiikan³⁷ tarkoitus on empiirisen havainnoinnin avulla selvittää, minkälaisia arvo- ja oikeudenmukaisuuskäsityksiä ja valtarakenteita tarkasteltavan järjestelmän parissa esiintyy. Se johdattaa aiheen kannalta merkittävien kysymysten äärelle. Analyttisen etiikan metodeilla voidaan puolestaan systematisoida ja problematisoida ilmiöiden keskeisiä ongelmakohtia ja havaita niissä uusia piirteitä. Muun muassa käsiteanalyysi mahdollistaa aiheen ydinkäsitteiden ja niiden keskinäisten yhteyksien teoreettisen tarkastelun, mikä syventää käytännön ongelman ymmärtämistä. Normatiivisen etiikan mukaisesti käytännöllisessä etiikassa pyritään rakentamaan systemaattisesti johdettuja toimintaehdotuksia ongelmiin.³⁸

Tekoälyn etiikka on usein muotoiltu periaate-etiikan mukaiseksi ratkaisukeskeiseksi toiminnaksi. Periaate-etiikan mukaiset ratkaisut tunnetaan yleisemmin keskitason periaatteina. Useat teknologiateollisuuden suuryritykset, kuten Google ja Microsoft, ovat julkaisseet yritysten sisäiset, tekoälyn kehitystä ohjaavat periaatteet. Mika Lintilä haastoi työ- ja elinkeinoministerinä toimiessaan suomalaiset yritykset seuraamaan monikansallisten yritysten esimerkkiä.³⁹ Montrealin yliopistossa kehitetty *Montreal Declaration for Responsible AI* (2017) on puolestaan hyvä esimerkki akateemisessa ympäristössä kehitetyistä keskitason periaatteisiin pohjautuvasta arvokoodistosta. Yritysten ja yliopistojen lisäksi tekoälyn etiikkaa

³³ Velvollisuus etiikan haaste on, että sen mukaan toimijan hyvä tarkoitus voi oikeuttaa tekoja, joilla on pahoja seurauksia.

³⁴ Seurausetiikkakin voi johtaa moraalisiin dilemmoihin, sillä pahoilla teoilla voidaan saada hyvää aikaan. Esimerkkinä lääkäri, joka tappaa ihmisen pelastaakseen useita ihmisiä elinsiirtojen avulla. Pietarinen 2015.

³⁵ Ollila 2019, 87.

³⁶ Kant 2004, 27; Pietarinen 2015; Ollila 2019, 47, 87.

³⁷ Deskriptiivinen etiikka saavutti 1900-luvulla suosiota muun muassa sosiologien, psykologien, biologisten ja kulttuuriantropologien keskuudessa. David Hume tutki moraalista empiirisen tutkimuksen kohteena jo 1700-luvulla.

³⁸ Airaksinen 1987, 21; Oksanen et al. 2010, 7–11; Hallamaa 2017, 8.

³⁹ Kauppinen & Laitinen 2018; Lintilä 2018; Microsoft AI Principles s.a.; Pichai 2018; Ollila 2019, 44.

on harjoitettu kansalaisjärjestöissä. *Future of Life* instituutti on yksi vaikutusvaltaisimmista tekoälyn kehitykseen keskittyvistä kansalaisjärjestöistä. Sen edustama työryhmä on kehittänyt *Asilomarin* periaatteiksi kutsutun koodiston.⁴⁰

Se, että keskitason periaatteet valikoituvat tekoälyn etiikan perustaksi, voi johtua useasta syystä. Ne ovat notkeita niin, että ne voidaan siirtää kontekstista toiseen. Ne voidaan muotoilla kattaviksi niin, että niihin voidaan sisällyttää monenlaisia arvoja, mikä näkyy siinä, että keskitason periaatteita ei yleensä esitellä kytkemällä ne johonkin analyttiseen kokonaisnäkemykseen, vaan ne voidaan helposti mukauttaa erilaisiin maailmankuviin. Keskitason periaatteet sopivat muodoltaan tekoälytutkimukseen, koska ne on melko helppo ilmaista yksinkertaisina ohjeina, jotka voidaan kirjoittaa koodiksi.⁴¹

Keskitason periaatteiden taustalla olevia arvoja ei yleensä selvitetä, kun periaatteet esitellään. Antti Kauppinen ja Arto Laitisen mukaan keskitason periaatteiden taustalle voidaan hahmottaa sellaiset yleiset arvot kuin (toimijan) autonomia, hyväntahtoisuus, ei-pahantahtoisuus, ihmiselämän itseisarvo, oikeudenmukaisuus ja ekologisuus.⁴²

Omien havaintojeni mukaan läheskään kaikki keskitason periaatteiden varaan rakentuvat mallit eivät kuitenkaan sisällä ekologisuuden vaatimusta, vaan Kauppinen ja Laitinen näyttävät lisänneen sen luetteloon, koska ajattelevat sen olevan tärkeä arvo. Esimerkiksi *Asilomarin* periaatteet perustuvat ihmiskeskeiseen näkökulmaan, vaikka ne ovatkin suhteellisen kattavat. Ollila ehdottaa, että *Asilomarin* periaatteet ja Montrealin yliopiston luomat periaatteet yhdistettäisiin yhdeksi kokonaisuudeksi, sillä Montreal Declaration for Responsible AI -koodistossa globaalin ekosysteemin kestävyys ja hyvinvointi on valittu toteutettavaksi tavoitteeksi.⁴³

Keskitason periaatteiden hyvä ja heikko puoli on niiden yksinkertaisuus. Näin voidaan muotoilla skaalautuva⁴⁴ eettisen ohjeistuksen perusmalli. Periaatteiden heikkous on puolestaan, että yksinkertaistus voi johtaa pitämään ongelmiakin yksinkertaisina ja estää kriittistä ajattelua. Mitä oikeudenmukaisuus käytännössä tarkoittaa? Kenen autonomiaa on suojeltava? Kysymykset osoittavat, että keskitason periaatteista voidaan johtaa hyvinkin

⁴⁰ *Asilomar Principles* 2017; *Montreal Declaration for AI* 2018.

⁴¹ Kauppinen & Laitinen 2018, 1–5; Ollila 2019, 44.

⁴² Keskitason periaatteita on hyödynnetty myös bioetiikassa. Kauppinen ja Laitilan mainitsema arvot vastaavat pitkälti bioetiikan keskeisiä arvoja. Bioetiikka on tekoälyn etiikkaa vanhempi tutkimusala ja voisi tarjota näkökulmia keskitason periaatteiden käytännön soveltamisesta. Beauchamp & Childress 2013; Kauppinen & Laitinen 2018, 1.

⁴³ Ollila 2019, 100.

⁴⁴ Skaalautuvuudella kuvataan yritystoiminnassa helposti laajennettavaa ja useassa kontekstissa hyödynnettävää toimintamallia.

erilaisia ja keskenään ristiriitaisia toimentasuosituksia. Maija-Riitta Ollila huomauttaa, että Kauppiisen ja Laitisen mainitsemat arvot edustavat länsimaista liberalismia ja ne edustavat lähinnä länsimaisten moraaliteorioiden näkemyksiä.⁴⁵ Periaatteet ohjaavat enemmänkin vastaamaan kysymykseen, kuinka eikä niinkään kysymyksiin miksi ja mitä. Tämän vuoksi keskitason periaatteet edellyttävät pohjakseen syvää analyysia tekoälyn keskeisistä ongelmakohdista ja moraalisisista dilemmoista. Tästä syystä ne eivät myöskään auta muodostamaan systemaattista arvopohjaa toiminnalle.

Tekoälyn etiikka on teknologian etiikan haara, joka on osa käytännöllistä etiikkaa. Täsmennän kuinka käytännöllinen etiikka ja soveltava etiikka eroavat toisistaan, sillä niitä käytetään usein synonyymeinä. Käytännöllinen etiikka on tutkimusta, jossa etiikan metodeja hyödynnetään käytännön ongelmien systemaattiseen tarkasteluun ja ratkaisuehdotuksien konstruomiseksi. Käytännöllinen etiikka on luonteeltaan preskriptiivistä. Soveltava etiikka on puolestaan käytännöllisen etiikan haara, jossa jotain normatiivisen etiikan teoriaa hyödynnetään käytännön ongelman ratkaisemiseksi. Kaikki soveltava etiikka on käytännöllistä etiikkaa, mutta kaikki käytännöllinen etiikka ei ole soveltavaa etiikkaa. Väitän, että ongelman konteksti voidaan ottaa paremmin huomioon, jos ongelmaa lähestytään ensin vain metodien avulla. Jos ongelmaa yrittää lähestyä valmiin moraaliteorian näkökulmasta, voi syntyä vinoumia, kun konteksti yritetään muokata teoriaan sopivaksi. Valmiin teorian taustalla on nimittäin paljon taustaoletuksia toimijuudesta, arvoista, oikeudenmukaisuudesta ynnä muista moraalifilosofian keskeisistä teemoista.⁴⁶ Deskriptiivisen, analyttisen ja preskriptiivisen etiikan vuorovaikutteisuus on tärkeää eettisten ongelmien tarkastelussa.

Teknologian etiikka tutkii usein teknologian kehitystä. Kehityksen käsite on arvottava: jotakin kehitetään huonommasta paremmaksi. Teknologian etiikassa tarkastellaan teknologioita, jotka vaikuttavat merkittävällä tavalla ympäristöön, jossa niitä käytetään ja joiden vaikutuksista tiedetään vähän tai niiden ajatellaan olevan eettisesti arveluttavia. Sotateknologiat, nanoteknologia ja robotiikka ovat hyviä esimerkkejä tällaisista tutkimuskohteista.⁴⁷ Tekoälyn etiikka on käytännöllisen etiikan tutkimusta, jossa pyritään hahmottamaan ja ratkaisemaan tekoälyn kehitystä ja hyödyntämistä koskevia eettisiä kysymyksiä.

⁴⁵ Ollila 2019, 45.

⁴⁶ Airaksinen 1987, 17, 21.

⁴⁷ Abney et al. 2011; Hansson 2017, 5; Ollila 2019, 70.

1.3 Vastuun tasot

Tekoälyn etiikassa on kolme termiä; teko, äly ja etiikka. Tässä yhteydessä etiikka kuvastaa käytännöllisen etiikan tutkimusta. Käytännöllisen etiikan tutkimuskohteena tekoälyä ei käsitellä vain teoreettisella tasolla, vaan sillä oletetaan olevan kausaalisuhde toimintaympäristöönsä, mikä vaikuttaa eettisen pohdinnan muotoihin ja painopisteisiin. *Teko* kuvastaa käsitteessä, että kyseessä on jokin keinotekoinen, mikä tarkoittaa, että ihminen on sen kehittänyt. Kyseisen termin merkitystä käsitteelle ei kovinkaan usein painoteta, vaikka käytännössä se on yksi kulmakivi tekoälyn moraalisisessa tarkastelussa.

Tietojenkäsittelytieteistä väitellyt Erkki Laitila tekee poikkeuksen kirjassaan *Ihanteeksi vastuullinen tekoäly* (2019). Hän kuvaa tekoälyn olevan ihmisten, tietokoneiden, verkostojen ja tieteen synergian tuotos, jolla pyritään kehittämään koko systeemin oppimiskykyä, tietoa ja ymmärrystä.⁴⁸ Laitilan määritelmässä tärkeä ja hyvin olennainen huomio on, että se korostaa ihmisen ja tekoälyn vuorovaikutteista suhdetta muita edellä kuvattuja määritelmiä vahvemmin. Tämä piirre auttaa muistamaan, että tekoäly ei synny ja leviä itsestään. Älykkyys ominaisuutena tarkoittaa sen haltijan kykyä autonomisuuteen ja oppimiseen eli adaptiivisuuteen. Kyseisten ominaisuuksien voidaan katsoa erottavan tekoälyn muista teknologioista.⁴⁹ Autonomisuudesta on tärkeää mainita, että sillä on eri asteita. Riippuu ohjelmoinnista, kuinka paljon tekoäly kykenee tekemään itsenäisiä päätöksiä. Autonomisuuden asteet korreloivat Boströmin ja Yudkowskin mainitsemien tekoälyn kehitystasojen kanssa. Mitä suurempi autonomian aste tekoälyteknologialla on, sitä vahvemman tekoäly on kyse.⁵⁰

Tekoälyn määritelmän mukaisesti aiheen muita tärkeitä tekijöitä ovat algoritmi, käytettävä alusta ja data. Ohjelmoijat kehittävät tekoälyssä hyödynnettävät algoritmit, jotka ohjaavat, kuinka tekoäly käsittelee ja käyttää informaatiota. Ohjelmoija luo koodiston, jolla tekoäly toimii.⁵¹ Tekoälyn etiikasta puhuttaessa viitataan monesti tieteiskirjailija Isaac Asimovin vuonna 1942 julkaisemassa *Roundout* teoksessa esitettyihin robotiikan kolmeen pääsääntöön.⁵² Säännöt ohjeistavat, kuinka robotti tulee ohjelmoida, jotta se ei vahingoittaisi ihmistä. Sääntöjä pidetään ensimmäisinä tekoälyn etiikan periaatteina, vaikka ne onkin

⁴⁸ Laitila 2019, 6.

⁴⁹ Ollila 2019, 53.

⁵⁰ Bostrom & Yudkowski 2014, 4.

⁵¹ Norvig & Russell 2010, 1039.

⁵² Ensimmäinen pääsääntö on, että Robotti ei saa vahingoittaa ihmisolentoa tai laiminlyönnin saattaa tätä vahingoittumaan. Toinen pääsääntö on, että robotin on noudatettava ihmisolentojen sille antamia sääntöjä, paitsi jos ne ovat ristiriidassa ensimmäisen pääsäännön kanssa. Kolmas pääsääntö on, että robotin on suojeltava omaa

kehitetty osaksi fiktiivistä tarinaa.⁵³ Ohjelmoijalla on vähintään osittainen vastuu tekoälyn toiminnasta.

Russell Ackoffin artikkelin *From data to wisdom* (1989) mukaan data on käsittelemättömiä symboleita, jotka heijastavat jonkin todellisuuden piirteitä. Ackoffin mukaan dataa täytyy prosessoida, jotta sitä voidaan hyödyntää informaationa. Näin ollen samasta datasta voidaan käsittelevästä riippuen saada erilaista informaatiota, mikä tarkoittaa, että informaatio ei ole objektiivista.⁵⁴ Datasta saatavan informaation sitä muokkaavan toimijan tavoitteita heijastavaa luonnetta voimistaa se, että tekoälyn käyttämä data rajataan tai päätetään olla rajaamatta tekoälyn käyttötarkoituksen mukaan. Rajaus sisältää oletuksen pätevästä datasta. Se, että data heijastaa joitakin todellisuuden piirteitä, asettaa datan valitsijalle myös vastuun selvittää, kuinka datan käyttö tekoälyssä vaikuttaa kyseiseen todellisuuteen. Kaikki tietokoneen hyödyntämä data on redusoitavissa binäärimuotoon, eli numeroista muotoutuvaksi koodistoksi.⁵⁵

Kuten kaikki teknologia, myös tekoäly kehitetään jotakin käyttötarkoitusta varten, kun joku haluaa saavuttaa hyötyä tekoälyä käyttämällä. Hyödyntäjä voi olla ohjelmoija itse, tekoälyn tilaaja, kuten yritys tai kunta tai kyseistä teknologiaa hyödyntävä kolmas osapuoli. Kehityksen ja hyödyntämisen käsitteiden avulla voimme havaita, että hyödyntäjällä on arvoja, joita hän pyrkii tekoälyn avulla saavuttamaan. Kehitys nimittäin sisältää asetelman, jossa kehitettävää asiaa tai asiantilaa parannellaan ja hyödyntäminen tähtää jonkin arvokkaaksi katsotun (hyödyn) saavuttamiseen. Hyödyntäjä myös muokkaa ympäristöään käyttäessään tekoälyä, mikä asettaa hänelle vähintään osittaisen vastuun muokkauksen lopputuloksesta.

Tekoälyn analyysissä erottuu kolme tärkeää aihetta; ohjelmointi, datan valinta ja käyttö sekä tekoälyn hyödyntäminen. Kaikkiin kolmeen aiheeseen vaikuttavat lisäksi tekoälyn toimintaympäristö ja käytettävä alusta, mitkä saattavat muokata tärkeiden kysymysten luonnetta.⁵⁶ Jaan tekoälyn etiikan kolmeen alakategoriaan seuraavasti: tekoälyn hyödyntämisen etiikka, tekoälyn ohjelmoinnin etiikka ja tekoälyn dataetiikka. Jaon on tarkoitus selvittää, minkälaisia vastuusuhteita tekoälyn kehityksessä vähintäänkin ilmenee. Sen ei ole tarkoitus tarjota tyhjentävää näkemystä, vaan helpottaa kysymysten

olemassaoloon, kuitenkin siten, että sen toiminnot eivät ole ristiriidassa ensimmäisen ja toisen pääsäännön kanssa. Asimov 1950.

⁵³ Tekoälyn etiikan kysymyksiä on muutoinkin käsitelty paljolti tieteisfiktion kautta, mikä on osiltaan vääristänyt keskustelun suuntaa. Murphey & Woods 2009.

⁵⁴ Ackoff 1999, 170–172.

⁵⁵ Norvig & Russell 2010, 53, 1039.

⁵⁶ Hyvä esimerkki on, että tekoälyä hyödyntävillä teollisuusroboteilla ja sosiaaliroboteilla on erilaiset toimintaympäristöt, jolloin myös niiden kohdalla painottuvat eettiset kysymyksetkin eroavat toisistaan.

muotoilemisessa. Vastuu on tärkeä käsite tekoälyn etiikassa, sillä tekoäly toimii ainakin osittain autonomisesti, mikä hämärtää vastuukysymyksiä. Mitä monimutkaisemmaksi tekoälyteknologia kehittyy, sitä vaikeammaksi vastuukysymyksetkin muuttuvat.

Tasot limittyvät ja ovat vuorovaikutuksessa keskenään. Tekoälyn etiikan keskusteluun osallistuvien onkin hyvä ymmärtää eri näkökulmien merkityksiä. Limittymisestä huolimatta jako helpottaa tärkeiden eettisten kysymysten havaitsemista. Ohjelmoinnin etiikan tärkeimpiä kysymyksiä on, kuinka ohjelmoida eettisesti toimiva tekoäly.⁵⁷ Huomattavaa on, että kysymys sisältää kaksi tasoa. Ensimmäinen taso on eettisen tekoälyn hahmottaminen (mitä tarkoittaa eettinen tekoäly?) ja toinen taso on etiikan ja tekoälyn yhdistäminen teknisenä suoritteena. Jaana Leikkaan ihmisslähtöisen teknologian teoria, jonka hän esittää kirjassaan *Life-Based Design - A holistic approach to designing human-technology interaction* (2009) on hyvä esimerkki ohjelmoinnin etiikasta. Leikas argumentoi, että teknologian taustalla vaikuttava ihmiskuva määrittelee teknologian kehityssuuntaa. Hän toteaa, että teknologian lähtökohtana tulee olla elämänlaadun parantaminen. Ennen teknologian kehittämistä tulee miettiä, mitä on laadukas elämä ja kuinka sitä voisi tukea.⁵⁸

Dataetiikassa pohditaan, minkälaisia vaikutuksia datan käytöllä voi olla tekoälyn toiminnalle tai datan heijastamalle todellisuudelle ja minkälainen data on ylipäänsä hyvää dataa. Mitä lisääntyvä datankeruu tarkoittaa yksilölle ja yhteisöille. Miten voidaan välttää datavinoumat⁵⁹ ja turvata datan turvallinen käyttö. EU:n tietosuoja-asetus GDPR on hyvä esimerkki dataetiikan käytännön ratkaisusta. GDPR:llä ei pyritä vain lisäämään yksilön suojaa ja itsemääräämisoikeutta digitalisaation aikakautena, vaan sillä luodaan myös raameja läpinäkyvämmälle datankeruulle ja käytölle.⁶⁰

Hyödyntämisen etiikka vastaa kysymyksiin kuten, mitä arvoja tekoälyn hyödyntämisellä halutaan tavoitella ja miten hyödyntäminen vaikuttaa yksilöön, yhteisöihin ja laajemmin yhteiskuntaan.⁶¹ Dataetiikka ja ohjelmoinnin etiikka ovat tästä näkökulmasta katsottuna alisteisia hyödyntämisen etiikalle. Valitut arvot ovat toimintaa ohjaavia päämääriä, joten datan käytön ja ohjelmoinnin tulisi olla linjassa hyödyntämisen etiikan kanssa.

⁵⁷ Kapealle tekoälylle voi olla helpompi ohjelmoida *käskyjä* jotka varmistavat esimerkiksi teknologian ihmisystävällisyyden. Yleisälyn kohdalla kysymys monimutkaistuu hyvinkin paljon, sillä kuinka ohjelmoida eettinen koodisto, kun on lähes mahdotonta ennustaa toiminnan kontekstia. Bostrom ja Yudkowsky 2011, 4.

⁵⁸ Leikas 2009, 13.

⁵⁹ Datavinouma kuvaa sitä, kun tekoälyn kehittämiseen ja päätöksentekoon valikoitu data sisältää ihmisten ennakkoluuloja tai päättelyvirheitä, jolloin ne toistuvat myös tekoälyn päätöksenteossa. Hern 2018; Ollila 2019, 120.

⁶⁰ Kauppinen & Laitila 2018, 6; Ollila 2019, 116; Fact Sheet: Artificial Intelligence for Europe s.a..

⁶¹ Ollila 2019, 81.

Ajatuksen voi myös kääntää toisin päin. Tarkkailemalla ohjelmoinnin etiikan ja dataetiikan käytännön toteutuksia, voidaan hahmottaa tekoälyllä tavoiteltavia arvoja. Hyödyntämisen näkökulma edustaa abstraktimpaa tasoa ja dataetiikka sekä ohjelmoinnin etiikka enemmän käytännön toteutuksen tasoa. Tutkielmani näkökulmana on tarkastella valtioita tekoälyn hyödyntäjinä, joten keskityn ensisijaisesti hyödyntämisen etiikan kysymyksiin. Tarkastelen seuraavassa luvussa, minkälaisia teemoja tekoälyn valtiollinen hyödyntäminen sisältää.

2 Arvot tekoälyn hyödyntämisen ohjaajina

Tekoälyn hyödyntämisellä on valtavasti potentiaalia hyvän luomiseen. Sen avulla voitaisiin kehittää nykyistä toimivampia tapoja vastata yhteiskuntien suurimpiin haasteisiin, kuten ilmastonmuutokseen, köyhyyteen ja väestörakenteiden muutoksiin.⁶² Haasteisiin vastaaminen vaatisi, että tekoälyn kehittäjillä olisi yhtenevät päämäärät toiminnalleen. Todellisuus on kuitenkin kaukana siitä. Tekoälyn yhteiskunnallista soveltamista tutkivan, tietojenkäsittelytieteistä väitelleen, Erkki Laitilan mukaan tekoälyn kehityksen hajautunut toiminta ja informaatio estävät demokratian, tasa-arvon ja tieteen kehittymisen.⁶³ Informaation ja toiminnan hajanaisuus johtuu siitä, että tekoälyä kehitetään suureksi osin tehostamaan yritysten voitontavoittelua. Valtioilla olisi tekoälyn hyödyntäjinä mahdollisuus tukea yhtenäisten päämäärien rakentumista. Hallitusten käytettävissä olevat mekanismit ovat esimerkiksi sidosryhmiä yhdistävien hankkeiden fasilitointi, lainsäädännön päivittäminen tekoälyaikakauden vaatimusten mukaiseksi,⁶⁴ rahoituksen ohjaus tai verohelpotusten jakaminen toivotun kaltaiseen innovaatiotoimintaan ja tekoälykehittäjien monimuotoisuuden vahvistaminen koulutuksen avulla.⁶⁵

Erkki Laitilan lisäksi valtiotieteiden tohtori Maija-Riitta Ollila ja Jerusalemin heprealaisen yliopiston historian professori Yuval Noah Harari ovat kirjoittaneet tekoälyn kehityksestä. Ollilan teos *Tekoälyn etiikkaa* (2019) keskittyy eettiseen pohdintaan, Hararin teos *Homo Deus, huomisen lyhyt historia* (2015) käsittelee aihetta historian tutkimuksen näkökulmasta ja Laitilan teos *Ihanteeksi vastuullinen tekoäly* (2019) pohjautuu systeemitieteisiin. Kirjoittajien teoksia yhdistää yhteiskunnallinen lähestymistapa ja ajatus, jonka mukaan tekoälyn kehityksen suunta on riippuvaisia tämän hetken arvokeskusteluista. Arvokeskustelun merkitys tekoälyn kehitykselle selittyy moraalifilosofian aksioomalla, jonka mukaan arvot ohjaavat toiminnan päämääriä. Toiminnan päämäärä rajaa ja ohjaa puolestaan toiminnon tapoja ja toivottuja/ei-toivottuja seurauksia.⁶⁶ Tarkastelen seuraavassa kahdessa osiossa Ollilan, Hararin ja Laitilan näkemyksiä tekoälyn yhteiskunnallisesta kehityksestä keskittyen tekoälykehityksen arvoviitekehykseen. Kolmannen osion lopuksi käsittelen *Creating Shared Value* (CSV) teoriaa tekoälyn etiikan kehittämisen työkaluna.

⁶² Chui et al. 2018.

⁶³ Laitila 2019, 26.

⁶⁴ EU:n tietosuojalaki GDPR on hyvä esimerkki tekoälyn vaikutuksia ennakoivasta lainsäädännöstä. Yleinen tietosuoja-asetus 2019.

⁶⁵ Chui et al. 2018, 31; Laitila 2019, 39; Ollila 2019, 343.

⁶⁶ Harari 2017; Laitila 2019; Ollila 2019, 93, 106.

Arvoista puhuvan täytyy täsmentää, mitä arvoja hän tarkoittaa välttyäkseen ekvivokaatiolta. Moraaliarvojen lisäksi on nimittäin olemassa subjektiiviseen käsitykseen pohjautuvia arvoja, kuten esteettisyyden, politiikan ja kulinaristisen nautinnon arvokäsitykset. Maija-Riitta Ollilan mukaan subjektiiviset arvot ovat moraaliarvoihin nähden välinearvoja, eli ihmiset toivovat saavuttavansa niiden avulla jotakin muuta. Moraaliarvot ovat subjektiivisista arvoista poiketen itseisarvoja, eli itsessään arvokkaina pidettäviä asioita. Moraaliarvot ovat luonteeltaan yleismaallisia. Ne edustavat sitä lopullista päämäärää, jota välinearvoilla tavoitellaan.⁶⁷ Etiikan historiaa käsittelevässä osiossa mainittu oikeudenmukaisuus on hyvä esimerkki moraaliarvosta. Pysin erottelemaan tekoälyn etiikan välinearvot ja moraaliarvot toisistaan.

2.1 Kehitystä kehityksen vuoksi

Yuval Hararin voidaan ymmärtää määrittelevän ihmisetkin algoritmeiksi. Harari todella sanoo kirjassaan, että ihmiset ovat algoritmeja, sillä heidän toimintansa on solutasolta lähtien geeneihin kirjattujen toimintaohjeiden noudattamista ja niiden ohjaamaa reagointia ympäristön luomiin paineisiin. Selityksen mukaan tiedostamaton ja luonnonvalinnan muokkaama solutason viestintä vaikuttavat tiedostetuiltakin vaikuttaviin päätöksiimme, kuten ostoslistan tekemiseen eikä toisin päin. Tietoisuus näyttäytyy selityksessä tarpeettomana miljardien neuronien välisen viestinnän muodostamana sivutuotteena.⁶⁸

Hararin kirjasta ei kuitenkaan kannata poimia yksittäisiä ajatuksia. Hänen tarkoituksensa on kuvata, millaisena maailma näyttäytyy, kun sitä tarkastellaan puhtaasti tieteellisen ja teknillisen kehityksen kautta. Harari kuvaa, kuinka 1900-luvun aikana tehty edistys biotieteissä on johtanut fysikaalisiin selityksiin niin ihmisen toiminnalle kuin ajattelullekin. Osa selittävistä tekijöistä on hermostossa kulkevia sähköimpulsseja, jotka aktivoivat kehon eri osa-alueita ja osa kemiallisia aineita, kuten adrenaliinia, tai endorfiinia, joita sähköimpulssit vapauttavat aiheuttaen ihmisessä tunnereaktioita. Sähköimpulssit ja niiden vapauttamat kemikaalit sisältävät dataa, joiden käsittely muodostaa ihmisen tunnetilat ja ajatusmaailman. DNA:n löytämisen myötä ihmisen yksilöllisiä ominaisuuksia on puolestaan voitu selittää geneettisen koodin avulla. Geeniperimä määrittää ennalta, minkälainen sähkökemiallinen informaation käsittelykokonaisuus ihmisestä muodostuu, elleivät satunnaiset mutaatiot vaikuta kehitykseen.⁶⁹

⁶⁷ Harari 2017, 329; Ollila 2019, 32.

⁶⁸ Harari 2017, 119, 329; Ollila 2019, 57.

⁶⁹ Harari 2017, 120, 298, 405.

Biotieteiden pohja perustuu evoluutioteorialle, jonka mukaan ihmisen geeniperimä on muotoutunut luonnonvalinnan kautta vuosituhansien aikana. Harari huomauttaa, että ihmisellä on nykyteknologian avulla kyky parantaa omaa evoluutiotaan. Biologia selittää ihmisen mekaniikan ja teknologia mahdollistaa sen kehittämisen. Biotekniikan avulla voidaan tarkkailla aivoalueen sähköimpulsseja ja tehdä havaintoja siitä, mitä ihminen ajattelee tai aikoo seuraavaksi tehdä. Biotekniikan ratkaisuilla kyetään jopa muokkaamaan ihmisen ajatustoimintaa syöttämällä aivoihin sähköimpulsseja päähän asetettujen elektrodien avulla.⁷⁰ Tekoäly on yksi tärkeimmistä biotekniikan kehityksen mahdollistavista teknologioista. Muun muassa ajatustoimintaa muokkaavassa teknologiassa algoritmit käsittelevät aivojen sähkötoiminnasta kerättyä suurta datamassaa ja ohjeistavat sen avulla, kuinka elektrodien tulisi syöttää sähköimpulsseja toivotun toiminnan saavuttamiseksi.⁷¹

Ajatustoimintaa muokkaavaa teknologiaa miettiessä tulee ajatelleeksi, mitkä kehityskulut ovat vaikuttaneet siihen, että tekoälyä hyödynnetään ihmisen keinotekoiseen paranteluun pyrkivän biotekniikan edistämiseen. Jos asiaa tarkastelee puhtaasti evoluution kautta, niin selitykseksi voi riittää, että biotekniikka on organismien kehitystä ja evoluutio johtaa kehittyvien organismien säilymiseen. Hararin mukaan asiaan vaikuttaa kuitenkin paljon monisyisempi polku, jonka hän tiivistää ajatukseen *ihmiskunnan asialistasta*. Se kuvastaa asioita, joita ihmiskunta on pyrkinyt teknologian avulla saavuttamaan lajin syntyhistoriasta asti. Asialista säilyi Hararin mukaan muuttumattomana 1900-luvun puoliväliin, mihin asti sen oli täyttänyt ulkoisten vaikuttajien määrittelemänä nälänhädän, kulkutautien ja sotien kukistaminen. Viimeisten viiden vuosikymmenen aikana ihmiskunnan kyky reagoida nälänhätään, epidemioihin ja sotien ehkäisemiseen on parantunut niin, että suurin osa ihmiskunnasta on vapautunut näistä vitsauksista. Asialistaan on muodostunut tyhjä kohta, jonka täyttäminen on tällä kertaa ihmisten käsissä.⁷²

Harari argumentoi, että liberalistisella maailmankuvalla, jonka mukaan ihmiskunta itsessään muodostaa merkityksen maailmankaikkeudelle, on suuri vaikutus asialistan uudelleenmuokkautumiseen. Ihmiskunnan itseisarvo perustuu vapaan tahdon teorialle, jonka mukaan vapaa tahto erottaa ihmisen muista elävistä organismeista. Vapaan tahdon ajatuksen myötä yksilön kokemukselle on annettu auktoriteetin asema tiedon määrittämisessä. Siitä on

⁷⁰ Kyseistä teknologiaa hyödynnetään myös muiden eläinten kognition tutkimiseen ja muokkaamiseen. Biotekniikan kehityksessä hyödynnetään erityisesti rottia koe-eläiminä. Harari 2017, 298.

⁷¹ Harari 2017, 294–297, 395; Koutonen 2018.

⁷² Harari 2017, 9, 73.

tullut jopa tieteellisesti tutkittua tietoa tärkeämpää, koska yksilön kokemuksesta johdetaan länsimaisen yhteiskunnan arvot.⁷³

Yksilön kokemukseen perustuva talousjärjestelmä on puolestaan muokkautunut kuluttajavetoiseksi. Hararin mukaan voidaan sanoa, että liberaalin talousjärjestelmän mukainen hyvä määräytyy sen mukaan, mikä tuottaa eniten. Sehän antaa tarkimman kuvan ihmisen kokemusmaailmasta. Hän huomauttaa, että liberaalin talousjärjestelmän mekanismit aiheuttavat varallisuuden epätasaista jakautumista, jossa pieni prosentti ihmiskunnasta hallitsee suurinta osaa kaikesta varallisuudesta. Näin ollen pienellä prosentilla ihmiskuntaa on mahdollisuus vaikuttaa liberalismien mukaisen hyvän määrittämiseen selvästi muuta ihmiskuntaa enemmän.⁷⁴

Hararin mukaan onnellisuus ja ihmiselämä ovat liberalismien perusarvoja. Biotekniikan myötä kuolema ja ihmisen kokema onnellisuus voidaan redusoida fysiologiseksi prosesseiksi, joita voidaan kehittää toivottuun suuntaan teknologian avulla. Kuolemaa voidaan pikkuhiljaa siirtää etäämmäksi ja yksilön sähkökemiallinen järjestelmä voidaan kehittää tuottamaan kokoaikaista onnellisuuden tunnetta.⁷⁵ Kuvailtu tilanne ei todennäköisesti ole kaikkien ihmisten toivoma tulevaisuuden kuva, mutta ei sen tarvitsekaan. Raha on demokratiaa suurempi auktoriteetti kehityskulun valinnassa. Hararin mukaan vapaan markkinajärjestelmän, liberalismien arvopohjan ja tieteen ja tekniikan yhdistetyt kehityskulut johtavat ihmiskunnan valitsemaan kuolemattomuuden, yksilöllisen onnellisuuden maksimoinnin ja jumalallisuuden uudeksi asialistakseen.⁷⁶ Suurten yksityisyritysten, kuten googlen tekemät investoinnit kuolemattomuuteen tähtääviin hankkeisiin tukevat huolestuttavasti Hararin esittämää visiota.⁷⁷

Kaiken edellä mainitun kehityksen taustalla vaikuttaa Hararin mukaan ensimmäinen kaikkia tieteitä yhdistävä paradigma: data-analyysi. Data-analyysi velvoittaa, että tutkimustieto on muutettavissa matemaattiseen muotoon keinotekoisten algoritmien⁷⁸ käsiteltäväksi. On parempi antaa keinotekoisten algoritmien muuttaa datamassat informaatioksi, sillä ne ovat siinä ihmistä tehokkaampia. Harari esittää huolensa siitä, että ihmiskunnan pyrkiessä yhtenäistämään tietojärjestelmänsä yhtenäisiksi kokonaisuuksiksi ja kehittäessään tekoälyteknologiaa tietomassojen käsittelyä varten, ihmisen yhteiskunnallinen

⁷³ Harari 2017, 73, 232, 293.

⁷⁴ Harari 2017, 39, 51, 356.

⁷⁵ Harari 2017, 48–50, 299.

⁷⁶ Harari 2017, 27, 63.

⁷⁷ Brooker 2015.

⁷⁸ Harari jaottelee eliöt orgaanisiksi algoritmeiksi ja tietotekniikan hyödyntämät algoritmit keinotekoisiksi.

rooli pelkistyy datan tuottajaksi.⁷⁹ Rooli sopisi biotieteiden mukaiseen ihmiskuvaan ja ajatukseen kehityksestä, mutta onko se ihmiskunnan kannalta toivottava tulevaisuus?

Hararin kuvaama länsimainen yhteiskunnallinen kehitys on linjassa kriittisen projektijohtamisen tutkimuksen kanssa. Projektijohtamisen tutkimuksen valtavirrasta poiketen, kriittinen projektijohtamisen tutkimus tähtää normatiivisten ohjeiden ja työkalujen rakentamisen sijaan kyseenalaistamaan alan keskeisiä taustaoletuksia ja kiinnittämään huomiota siihen, kuinka projektijohtamisen vaikutukset näkyvät yhteiskunnallisella tasolla. Kriittistä tutkimusta tekevien tutkijoiden pyrkimyksenä on uudistaa projektijohtamisen tutkimus yhteiskunnan kannalta merkityksellisemmäksi.⁸⁰

Bristolin yliopiston kauppakorkeakoulun professori Svetlana Cicmil ja Manchesterin yliopiston organisaatioanalytiikan professori Damian Hodgson sanovat artikkelissaan *New Possibilities for Project Management Theory: A Critical Engagement* (2006), että projektijohtaminen kehitettiin 50-luvulla ratkaisemaan muun muassa teknologiateollisuutta vaivannutta tehottomuuden ja korkeiden kustannusten ongelmaa, ja selkeyttämään organisaatioiden monimutkaistuvaa toimintaympäristöä.⁸¹ Tehottomuuden ja korkeiden kustannusten ajateltiin johtuvan ihmisluonnon laiskuudesta ja taipumuksesta vilppiin. Ulkoista kontrollia lisäämällä uskottiin voitavan ratkaista tällaiset ongelmat. Työnteon mitattavuudesta tuli tärkeä osa kontrollia ja projektijohtamista.⁸²

Cicmilin ja Hodgsonin mukaan projektioirganoitumisen mallin ajateltiin tehostavan työntekoa, ja siitä tuli nopeasti työelämän uusi trendi, mikä johti yhteiskunnan projektisoitumiseen. Kirjoittajat tarkoittavat projektisoitumisella sitä, että projektijohtamisen ihanteet kuten instrumentalismi, tiedon universalisuus, tehokkuus, mitattavuus ja vaikuttavuus kasvattavat merkitystään myös työelämän ulkopuolella.⁸³

Instrumentalismi tarkoittaa etiikan näkökulmasta sitä, että asioita arvotetaan niiden käytettävyyden mukaan. Mitä paremmin teoria, toimintatapa tai ystävyysuhteet tuottavat käytännön hyötyä, sitä arvokkaampia ne ovat. Tiedon universalisuus tarkoittaa tiedon yleistettävyyttä kontekstista toiseen. Cicmilin ja Hodgsonin kuvauksessa se muun muassa

⁷⁹ Harari 2017, 386, 404.

⁸⁰ Cicmil ja Hodgson 2006, 3; Oksman 2013, 24–25.

⁸¹ VUCA:n käsitettä käytetään monimutkaistuvan ympäristön selittämiseksi. Käsite on lyhenne sanoista epävakaa (Volatile), epävarma (Uncertain), kompleksisuus (Complexity) ja monitulkintaisuus (Ambiguity). Käsitteen mukaan toimintaympäristöä ei voida hallita. Nathan Bennetti ja James Lemoine mukaan VUCA:n kaltaisessa ympäristössä voidaan toimia, jos VUCA:n piirteet otetaan huomioon toiminnan suunnittelussa. Suunnittelulla varmistetaan toimijoiden reaktio- ja adaptaatiokykyjä. Bennett ja Lemoine 2014.

⁸² Cicmil & Hodgson 2006, 14–18.

⁸³ Cicmil & Hodgson 2006, 14–16.

tarkoittaa sitä, että se, mikä todetaan yhdessä projektissa toimivaksi ratkaisuksi, oletetaan toimivan kaikkien projektien kohdalla. Yleistettävyyden oletus vähentää sosiaalisen ja fyysisen ympäristön huomioimista ja merkitystä osana inhimillistä toimintaa.⁸⁴

Kirjoittajat tarkoittavat tehokkuuden, mitattavuuden ja vaikuttavuuden ihannoinnilla sitä, että toimien todellisilla seurauksilla ei ole suurta painoarvoa, kunhan niillä on teoriassa mahdollisuus vaikuttaa inhimillisen toiminnan tehostumiseen, ovat luonteeltaan mitattavia ja sitä kautta niillä voidaan osoittaa, että toiminnalla on vaikuttavuutta.⁸⁵ Tällaisessa tilassa ei kiinnitetä kovinkaan paljoa huomiota siihen, mitä ovat ne perimmäiset asiat, joita inhimillisellä toiminnalla on tarkoitus saavuttaa. Kunhan se, mitä tehdään, tehdään tehokkaasti. Ilmiö tuottaa näennäistä tehokkuutta ja vaikuttavuutta.

Tehokkuus, mitattavuus ja vaikuttavuus ovat inhimillistä yhteistoimintaa tukevia välinearvoja, joiden on tarkoitus auttaa varsinaisten tavoitteiden saavuttamista. Cicmilin ja Hodgsonin mukaan yhteiskunnan projektisoitumisen myötä ne ovat kuitenkin vääristyneet itseisarvon asemaan. Instrumentalismi ja objektiiviseen tietoon uskomisen edustavat puolestaan maailmankatsomusta, jolla ihmisille syntyy valheellinen kontrollin tunne omasta ympäristöstään. Projektijohtamisen parissa vallitseva ajatus parhaiden käytäntöjen (best practice) universaaliuudesta on erinomainen esimerkki objektiivisen tiedon oletuksesta.⁸⁶

Cicmilin ja Hodgsonin tapaan Tukholman kuninkaallisen teknillisen korkeakoulun teollisen taloustieteen ja johtamislaitoksen professori Johan Packendorff on tutkinut länsimaisten yhteiskuntien projektisoitumista. Hän kuvaa artikkelissaan *The temporary society and its enemies: Projects from an individual perspective* (2002), kuinka projektisoituminen johtaa ihmiskuvan kaventumiseen, kun instrumentalistinen ajatusmalli heijastuu kaikille elämän alueille. Perhe, ystävyysuhteet ja vapaa-aika joutuvat muokkautumaan jaksottaiseen ja työntekijältä joustoa⁸⁷ vaativaan työrytmiin sopiviksi. Samoin vapaa-ajan aktiviteetteja pyritään määrittämään ja arvottamaan teknis-rationaalisen maailmankuvan mukaisesti toiminnan tehokkuuden ja vaikuttavuuden mukaan.⁸⁸

Edellä esitetyt yhteiskunnallisen kehityksen visiot saavat aikaan epämukavan tunteen, etenkin Hararin dystooppinen kuvaus ihmiskunnan asialistasta. Epämukavuus johtuu siitä, että visioissa ei ole sijaa inhimillisyydelle. Kirjoittajat pyrkivät epämukavuuteen tähtäävällä retoriikalla vapauttamaan lukijansa ajatuksesta, että teknologiaan ja niiden kehitykseen

⁸⁴ Cicmil & Hodgson 2006, 3–4, 14.

⁸⁵ Cicmil & Hodgson 2006, 14–18.

⁸⁶ Cicmil & Hodgson 2006, 3–4, 14.

⁸⁷ Jousto tarkoittaa tässä kohden erityisesti vaihtelevia työvuorojen pituuksia.

vaikuttaviin yhteiskunnallisiin kehityssuuntiin tulisi suhtautua neutraalisti. Kun nykytila problematisoidaan, on helpompi pohtia, minkälainen kehitys olisi toivottavaa. Harari sanoo hyvin kirjassaan, että historian ymmärtäminen vapauttaa parhaassa tapauksessa ihmisen ajattelemaan tulevaisuutta ilman, että on kahlittuna menneeseen.⁸⁹

2.2 Systeemievoluutio

Evoluutio ei voi lakata, mutta sitä voidaan ohjata. Tämä lause on monen teknologista kehitystä edistävän tahon argumentti, jolla he perustelevat näkemyksiään tarvittavasta kehityssuunnasta. Nick Bostrom hyödyntää sitä perustellessaan, miksi ihmisten tulisi pyrkiä kohti postumaania aikakautta⁹⁰. Ray Kurzweil käyttää sitä argumentoidessaan, miksi ihmisten tulisi pyrkiä sulautumaan yhteen tietokoneiden kanssa. Erkki Laitila käyttää sitä perustellakseen, miksi yhteiskunnallinen järjestelmä pitäisi rakentaa tekoälyvetoiseksi.⁹¹ Lause on myös todenmukainen. Ekosysteemien muuttuessa myös elävien organismien on muututtava tai ne katoavat. Hararin esittämät teknologiat osoittavat, että ihminen kykenee keinoitekoisesti vaikuttamaan ympäristönsä ja biologiansa kehitykseen.⁹² On kuitenkin tärkeää ymmärtää, että vaikka lause on todenmukainen, niin sen jälkeen esitetyt kehityssuunnat eivät ole evoluution välttämättömiä tuotoksia, vaan mahdollisia vaihtoehtoja tuhansien tai miljoonien vaihtoehtojen joukossa. Evoluutioon vetoaminen on näin ollen retorinen keino mahdollisesti arveluttavina pidettävien tai vähän tunnettujen teorioiden legitimoimiseksi eikä osoitus teorian faktuaalisesta perustasta.

Muun muassa posthumanismi perustuu ajatukselle, että ihmiskunta on jo nyt lähellä teknologista singulariteettia, millä tarkoitetaan tilannetta, jossa supertekoäly ylittää ihmisen älyllisen kapasiteetin ja teknologiakehitys muuttuu ennustamattomaksi. Posthumanismin edustajien mukaan ainoa tapa, jolla ihminen voi säilyttää merkityksellisyytensä singulariteetin jälkeisessä maailmassa, on kehittää omia kognitiivisia, fyysisiä ja emotionaalisia kykyjään yli-inhimilliselle tasolle teknologian avulla. Tietojenkäsittelytieteiden tutkijat ovat kuitenkin eri mieltä siitä, onko teknologista singulariteettia edes mahdollista saavuttaa.⁹³ Mielestäni olisi ennemmin syytä miettiä, minkälaiset kehityssuunnat ovat toivottuja kuin tehdä valintoja kiireen tunteessa, ja herätä tehtyjen valintojen heikkouteen ja peruuttamattomuuteen.

⁸⁸ Packendorff 2002.

⁸⁹ Harari 2017, 67.

⁹⁰ Bostrom kuvaa, että postumaani aikakausi alkaa hetkestä, kun ihminen on kehittänyt jotakin fyysistä ja/tai psyykkistä ominaisuuttaan yli-inhimilliselle tasolle teknologian avulla.

⁹¹ Kurzweil 2001; Bostrom & Sandberg 2008; Bostrom 2009; Laitila 2019, 69.

⁹² Harari 2017, 294.

⁹³ Laitila 2019, 60–62.

Evoluutio ei välttämättä tarkoita pelkästään elävien organismien kehitystä, vaan sitä voidaan hyödyntää kuvaamaan myös systeemien kehitystä. Erkki Laitila puhuukin evoluutiosta nimenomaan systeemien välisenä kamppailuna eloonjäämisestä. Laitila rakentaa *Ihanteeksi vastuullinen tekoäly* -kirjassaan teoriapohjaa tekoälyn yhteiskunnalliselle soveltamiselle.⁹⁴ Laitilan teos on ansiokas siinä, että se mahdollistaa tekoälyn yhteiskunnallisen hyödyntämisen tarkastelun *holarkian* käsitteen kautta, joka on hänen kuvaamansa systeemiteorian kulmakivi. Käsite yhdistää holismin ja hierarkian ajatukset yhdeksi kokonaisuudeksi. Holismin tapaan holarkiassa ajatellaan, että systeemin osien välinen vuorovaikutus muuttaa kokonaisuuden emergentiksi niin, että vuorovaikutuksen luomaa lisäarvoa ei voida redusoida osiensa summaksi. Holarkian osat voidaan kuitenkin asettaa hierarkkiseen suhteeseen sen mukaan, miten ne vaikuttavat systeemin kokonaisuuteen.⁹⁵

Laitila määrittelee tekoälyn ihmisten ja teknologian väliseksi vuorovaikutteiseksi suhteeksi, joka pyrkii systeemitasoisen ymmärryksen kehittämiseen. Puhuessaan tutkielmani määritelmän mukaisesta itse teknologiaa tarkoittavasta tekoälystä Laitila käyttää keinoälyn termiä. Hänen mielestään yhteiskunnat voisivat kehittyä keinoälyn avulla systeemievoluution seuraavalle tasolle muuttamalla palveluiden tarjonnan ja poliittisen päätöksenteon *työntöohjauksesta imuvetoiseksi*. Työntöohjaus kuvastaa ylhäältäpäin määriteltyjen tarpeiden muodostamaa päätöksentekoprosessia ja imuvetoisuus puolestaan siirtymistä ympäristöstä ja ihmisten tarpeista kerätyn tiedon⁹⁶ mukaiseen päätöksentekoon.⁹⁷ Imuvetoinen päätöksentekoprosessi painottaa holarkian mukaisesti yhteiskunnan osien välistä vuorovaikutuksen merkitystä.

Systeemitieteiden mukainen käsitys holarkiasta perustuu moduuliajattelulle. Laitilan mukaan moduuliajattelu on yksi keskeisimmistä tekijöistä, kun tekoälyä kehitetään yhteiskunnallista soveltamista varten. Moduulit kuvastavat monimutkaisista ja laajoista

⁹⁴ Laitilan teoksessa on paljon puutteita akateemisesti tarkasteltuna, mutta jos sitä käsittelee keskustelunavauksena, ei valmiin teorian esittäjänä, teos tarjoaa hyviä näkökulmia tekoälyn yhteiskunnalliseen hyödyntämiseen ja kehittämiseen. Akateemisia puutteita mainitakseni, Laitila viittaa usein asiantuntija-auktoriteettina Metayliopistoon, joka pienellä tarkastelulla paljastuu Laitilan itsensä perustamaksi foorumiksi. Hän jakaa foorumissa ajatuksiaan tekoälyn kehityksestä. Laitila myös esittäytyy useiden tieteidenalojen asiantuntijana ilman kyseisten alojen koulutusta tai päteviä lähteitä tukemaan ajatuksiaan. Hän muun muassa sanoo, että nykypäivän valtiotieteissä ei tehdä laadukasta tutkimusta, koska se ei tuota selkeään muotoon saatavaa metatietoa tutkimistaan ilmiöistä. Selkeä muoto tarkoittaa Laitilalle koodiksi sopivaa tietoa, mikä palauttaa ajatukset Hararin ajatukseen data-analytiikasta kaikkia tieteitä yhdistävänä paradigmatena. Laitila 2019, 6, 28, 69.

⁹⁵ Laitila 2019, 7, 36; Kokonaisvaltainen ja kestävä systeeminen muutos s.a..

⁹⁶ Tietoa edeltävät asteet ovat informaatio ja data. Keinoäly olisi Laitilan mukaan erinomainen väline jalostamaan kerätty data informaatioksi ja tiedoksi, joiden perusteella yhteiskuntaa ohjaavat tahot voisivat tehdä päätöksiä.

⁹⁷ Laitila 2019, 94, 106.

tietomääristä kehitettyjä metatasoja.⁹⁸ Yksinkertaistavana esimerkkinä mäntää, bensansyöttömekanismia ja sytytinjärjestelmää yhdistää se, että ne ovat kaikki bensiinimoottorin osia. Niin kuin moottori muodostaa muiden laitteiden, kuten korin, akun ja ohjausjärjestelmän kanssa kokonaisuuden, jota kutsutaan autoksi, myös moduulit voivat keskenään muodostaa laajempia kokonaisuuksia. Autoesimerkki voidaan laajentaa käsittämään useiden autojen muodostama tieliikennejärjestelmä, joka puolestaan voidaan nähdä osana maailmanlaajuista logistiikkajärjestelmää. Näin myös moduuleja on mahdollista yhdistää aina vain suuremmiksi kokonaisuuksiksi, joita metatieto yhdistää. Logistiikkajärjestelmä yhdistettynä koulutusjärjestelmään, talousjärjestelmään, valtion hallintoon, kulttuurituotantoon, sosiaalipalveluihin ynnä muihin monimutkaisiin kokonaisuuksiin muodostaa puolestaan yhteiskunnallisen toiminnan raameja. Moduuliajattelun samanaikainen vahvuus ja heikkous on, että se yksinkertaistaa monimutkaisia kokonaisuuksia. Sen vuoksi Laitila ajattelee, että tekoälyä ei ”rekrytoida työhön”⁹⁹, vaan sillä rakennetaan yhteistyömalleja. Laitila tarkoittaa tällä, että tekoälyn ei tulisi olla yhteiskunnallisten päätösten ratkaisija vaan ratkaisua avustava työkalu.

Laitilan mukaan kehittämällä keinoälyä holarkian periaatteiden mukaisesti, keinoälyn toiminta ja vaikutukset voidaan saada läpinäkyviksi. Läpinäkyvyys toteutuu, kun moduuleista tehdään rekursiivisia niin, että ne voidaan redusoida pienempiin osiinsa aina ensimmäisen moduulin osiin asti. Tämä helpottaa tiedon muodostumisen seuraamista, kun sen koostumisen eri vaiheet voidaan saada näkyviin.

Vaikka tiedon muodostamisen virhemarginaalia voitaisiin seurata yksittäisten moduulien tarkkuudella, kokonaisuuden toimivuutta ei voida parantaa keskittymällä yksittäisten osien paranteluun. Pyrkimystä parantaa systeemiä yksittäisten osien kautta kutsutaan osaoptimoinniksi. Osaoptimointi ei ota systeemin vuorovaikutteista tasapainotilaa huomioon, minkä vuoksi sillä on todennäköisemmin negatiiviset vaikutukset osien väliseen vuorovaikutukseen kuin positiiviset.¹⁰⁰ Esimerkiksi liikennejärjestelmän turvallisuutta ei voida parantaa vain asettamalla teille rajoitteita. Autoa ajavien ihmisten on ymmärrettävä liikenteen järjestelyjä ja ymmärrettävä vastuunsa osana liikenneturvallisuuden muotoutumista. Liikennejärjestelyiden ja ihmisten toimintatapojen kohdatessa toisensa, saavutetaan liikenneturvallisuudessa huomattavasti selvempi kohennus kuin pelkkien rajoitteiden lisäämisellä.

⁹⁸ Laitila 2019, 16, 90.

⁹⁹ Laitila 2019, 60.

Tekoälyn vastuullisuudella Laitila tarkoittaa toimivaa ja läpinäkyvää systeemiä. Tekoälyn modulaarinen kehitys varmistaisi läpinäkyvyyden. Toimivalla systeemillä Laitila tarkoittaa puolestaan parhaiksi havaittujen arvojen toteuttamiseen pyrkivää ja rakenteiltaan kestävästä kokonaisuudesta. Laitila ymmärtää arvot kontekstisidonnaisina ja ajattelee, että niiden määrittelyssä tulisi kuunnella yhteisön kaikkia toimijoita. Laitilan määrittelemä vastuullisuus sisältää ajatuksen aiemmin mainitusta imuvetoisesta yhteiskunnallisesta ohjauksesta. Imuvetoinen ohjaus määrittelee yhteiskunnan tarpeet ja arvot sisältäpäin, jolloin yhteiskunta, eli sen muodostamat kansalaiset siirtyvät tekoälykehityksen objektista toimintaa määritteleväksi subjektiksi.¹⁰¹

2.4 Ihmisen ja tekoälyn yhteys

Aseet eivät tapa ihmisiä, ihmiset tappavat ihmisiä. Tämä tunnettu lause osoittaa, että jopa ilmeisen arveluttavia teknologioita pidetään moraalisesti neutraaleina, ja että moraalinen tarkastelu on relevanttia vain suhteessa teknologioiden käyttäjiin. Yhdysvalloissa tämän tapaisella argumentoinnilla vaikutetaan jopa aseita rajoittavien lainsäädäntöjen asettamiseen.¹⁰² Maija-Riitta Ollila suhtautuu Hararin tapaan kriittisesti tämän kaltaisiin käsityksiin teknologian neutraalisuudesta. Hän problematisoi ajatusta käyttämällä affordanssin käsitettä. Affordanssi tarkoittaa teknologian ominaisuutta toiminnan mahdollistajana. Käsitteen esitti ensiksi psykologi James J. Gibson (1904-1979).¹⁰³ Ihminen on rakentanut teknologiat, kuten muutkin artefaktit, jottain käyttötarkoitusta varten. Näin ne eroavat epäorgaanisista luonnonmateriaaleista, kuten kivistä ja puista.¹⁰⁴ Käyttötarkoitus antaa teknologian toiminnalle päämäärän, joka ilmaisee arvon, jota sen tulisi toteuttaa. Teknologian suunnitellut henkilö ei kuitenkaan voi määritellä affordanssia, vaan se määräytyy kunkin käyttäjän toiminnassa.¹⁰⁵ Yksi vuotias lapsi voi nähdä matkapuhelimessa imeskelyyn soveltuvan laitteen, 15-vuotias nuori voi ajatella sen olevan identiteetin rakentamiseen ja itseilmaisuuksiin soveltuva väline ja koira voi nähdä sen hyvänä puruluun korvikkeena.

Affordanssin käsitteen avulla teknologioita voidaan tarkastella moraalifilosofisesti sillä perusteella, että ne ovat olemassa inhimillisen tavoitteellisen toiminnan ansiosta. Moraalisen intensiteetin käsite voi auttaa ohjaamaan moraalisen tarkastelun kohteeksi otettavien

¹⁰⁰ Laitila 2019, 36–39, 114.

¹⁰¹ Laitila 2019, 29, 50.

¹⁰² Kainiemi 2019.

¹⁰³ Gibson 1979.

¹⁰⁴ Kivillä ja puillakin voidaan sanoa olevan affordanssin ominaisuus, mutta ihminen ei ole vaikuttanut niiden kehittymiseen.

¹⁰⁵ Gibson 1979; Ollila 2019, 64.

teknologioiden valintaa. Käsitteen mukaan teknologian muodostamien ongelmien haastavuus, vaikutusten ajankohtaisuus ja laajuus vaikuttavat siihen, miten merkittävä tarkasteltava teknologia on moraalinen näkökulmasta.¹⁰⁶

Tekoälyyn yleisesti yhdistettyjä affordansseja ovat laskennallisen toiminnan tehostaminen, epäorgaanisten asioiden automatisoiminen ja adaptiivisuus. Kapeakin tekoäly kykenee käsittelemään oikeaan muotoon muutettua dataa miljoonia kertoja nopeammin kuin ihminen ja ohjaamaan autoa sujuvasti liikenteessä ilman ihmisen vaikutusta.¹⁰⁷ Lisäksi tekoäly sisältää yleishyödyllisyyden affordanssin, sen on yleishyödyllinen teknologia (*general purpose technology*). Yleishyödyllistä teknologiaa voidaan hyödyntää laaja-alaisesti eri teollisuuksien välillä. Tekoälyn osalla hyödynnettävyyden rajaa ei näytä olevan. Yleishyödylliset teknologiat ovat puolestaan ominaisuuksiltaan disruptiivisia.¹⁰⁸

Disruptiivisuus tarkoittaa teknologioiden kontekstissa luovaa tuhoa, jossa uuden teknologian mahdollistama kehitys tekee vanhoista teknologioista ja niihin liittyvistä toimintatavoista tarpeettomia. Uusi teknologia mahdollistaa, että poistuneiden toimintatapojen ja työpaikkojen tilalle muodostuu jotakin uutta, ennustamatonta. Sähköteknologian, internetin ja mobiiliteknologian kehitys ovat hyvä osoitus yleishyödyllisten teknologioiden vaikutuksista. Ne ovat olleet keskeisiä tekijöitä maailman digitalisoitumisessa. Koska tekoälyn hyödyntämisellä ei näytä olevan alakohtaisia rajoja, sen tuottama luova tuho tulee todennäköisesti olemaan laajempaa kuin minkään aikaisemman teknologian.¹⁰⁹

Luova tuho tarkoittaa organisaatioiden osalta, että muutoksen jalkoihin jääneiden organisaatioiden ja työpaikkojen muodostamaa tyhjiötä tulee täyttämään uudet ja uusiutuneet organisaatiot ja työskentelyn muodot. Ray Kurtzweilin, Googlen koneoppimisesta vastaavan jaoston johtajan mukaan teknologian aiheuttamaa luovaa tuhoa ei pidä pyrkiä hallitsemaan, sillä sen lopputuloksena syntyvää arvoa ei voida ymmärtää ennalta.¹¹⁰ Näkemys antaa kuvan, että teknologinen kehitys on ihmisen ulottumattomissa tapahtuva ilmiö, vaikka nimenomaan ihminen kehittää toiminnallaan tekoäly ja muita yleishyödyllisiä teknologioita ja levittää niitä. Taloustieteessä hyödynnetyn näkymättömän käden –mekanismin kaltaisen aktantin hahmottaminen luovan tuhon ohjaajaksi vaikuttaa enemmän tietoiselta pyrkimykseltä väistää eettisesti haastavien valintojen pohdinta tai omien toiveiden oikeuttamiselta kuin todelliseen

¹⁰⁶ Jones 1991; Ollila 2019, 34.

¹⁰⁷ Kurtzweil 2001; Bostrom & Yudkowsky 2014; Ollila 2019, 53.

¹⁰⁸ Jovanovic & Rousseau 2005.

¹⁰⁹ Jovanovic & Rousseau 2005; Norvig & Russell 2010, 1034; Harari 2017, 328.

¹¹⁰ Kurtzweil 2001; Jovanovic & Rousseau 2005; Ray Kurtzweil Biography 2019.

tilanteeseen pohjautuvalta argumentilta. Se, että jokin asia on äärimmäisen vaikea saada aikaan, ei tarkoita, ettei sitä voitaisi saada aikaan.

Tekoälyn disruptiivinen luonne saa kysymään, onko joidenkuiden kärsittävä, jotta jokin uusi voi kukoistaa, vai voisiko luovan tuhon positiivisia vaikutuksia nopeuttaa ja kärsimyksen määrää minimoida. Peter Norvig ja Stuart Russel mainitsevat nopeat työelämämuutokset mahdollisina inhimillisen kärsimyksen aiheuttajina, kun puhutaan tekoälyn yhteiskunnallisista vaikutuksista. Heidän mukaansa työelämän muutosta olisi tutkittava niin, että tekoälyn automatisoimien töiden aiheuttamaa inhimillisen merkityksellisyyden tunteen katoa voitaisiin lieventää ja työhön perustuvaa varallisuuden jakoa korvata tai uudistaa.¹¹¹

Norvigin ja Russellin ajatukset tekoälyn yhteiskunnallisista vaikutuksista koskevat lähitulevaisuutta. Yuval Hararin argumentit tekoälykehityksen vaikutuksista ovat kauaskantoisempia. Hänen huolensa koskee ihmiskunnan kehitystä kaikessa laajuudessaan, ja teknologinen kehitys on sen keskiössä. Hararin kuvaus ihmisen ja tekoälyn välisestä suhteesta on, että ihmisen rooli on pikkuhiljaa kaventumassa tekoälyn tarvitseman tiedon tuottajaksi. Kyseistä kehitystä ei ole Hararin mukaan helppoa hahmottaa, sillä teknologia on ollut luonnollinen tuki ihmiskunnan kehityksessä. Ei ole helppoa ajatella, että roolit olisivat vaihtumassa, ja ihmisestä olisi muodostumassa tuki teknologian kehitykselle.¹¹²

Edellä kuvatuissa visioissa tekoäly vaikuttaa ihmisistä riippumattomalta autonomiselta toimijalta. Koen, että kyseisten visioiden haitallinen puoli on, että ne etäännyttävät tekoälyn kehityksen inhimillisestä toiminnasta. Laitila kuvaa teknologian ja ihmisen vuorovaikutteista suhdetta sanalla *sosio-tekninen*. Sosio-teknisen käsite hahmottaa teknologisen kehityksen raameja niin, että teknologian tulee sopia ihmisen toimintaympäristöön, jotta sen käyttö leviää. Laitilan mukaan tekoälyn holarkinen kehitys edellyttää, että tekoälyä tarkastellaan sosio-teknisten suhteiden osapuolena. Yhteiskunnallisesta näkökulmasta se tarkoittaa, että kehitystä ohjaavien tahojen tulisi osallistaa kansalaiset ja yhteisöt teknologian kehitystä määrittäviksi subjekteiksi. Nykykyisessä yritystoiminnan määrittämässä kehitystyössä yhteisöt ja kansalaiset jäävät ennemmin kehityksen objekteiksi.¹¹³

2.5 Markkinoiden voimalla päin seinää

Hararin, Ollilan ja Laitilan teoksia yhdistää näkemys, jonka mukaan tekoälyn kehitystä ei tulisi jättää markkinavoimien hoidettavaksi. Kyseinen näkemys johtuu siitä, että

¹¹¹ Norvig & Russell 2010, 1034.

¹¹² Harari 2017, 56, 382.

¹¹³ Laitila 2019, 26; Tieteen termipankki 2019.

markkinavetoisuus muodostaa kapea-alaista kehitystä, jossa moraaliset vinoumat voidaan hyväksyä lyhyen aikavälin taloudellisen voiton maksimoimiseksi. Hyvä esimerkki kapea-alaisesta kehityksestä on, että tekoälyn kehityksen suurin kohde yksityisellä sektorilla rahallisesti mitattuna on mainonnan kohdentaminen.¹¹⁴ Mainonnan kohdentamisella on tarkoitus tehostaa kulutustuotteiden ja palveluiden tuottavuutta. Näin ollen tekoälyn hyödyntämisen perimmäisen merkityksen määrittelemisen markkinoiden mukaisesti ohjaisi tekoälyn potentiaalin palvelemaan kulutusyhteiskuntaa. Näin tarkasteltuna kansalaisten rooli kaventuu kuluttajiksi ja yhteiskunta tekoälyn hyödyntämisen objektiksi.

Edellä kuvattu tilanne on erinomainen esimerkki markkinahäiriöstä. Markkinahäiriöllä tarkoitetaan tilannetta, jossa vapaakauppa ja vapaat markkinat eivät johdakaan palveluiden ja tuotteiden osalta ihmisten toivomaan lopputulokseen.¹¹⁵ Tekoälyn avulla voisi olla mahdollista löytää nykyistä tehokkaampia keinoja hillitä ilmastonmuutosta ja hallinnoida yhteiskunnan resursseja ihmisten käytännön tarpeet ja toiveet huomioiden. Tästä näkökulmasta katsottuna harva sanoisi, että kohdennettu mainonta on tekoälyteknologian hyödyntämistä parhaimmillaan.

Ollilan mukaan yksi syy markkinahäiriöihin on, että *pukki vahtii kaalimaata*.¹¹⁶ Ollila tarkoittaa analogialla sitä, että jos eettinen säätely jätetään yrityksille itselleen, he todennäköisesti hyödyntävät sitä ajaakseen omia etujaan ennemmin kuin yhteistä hyvää. Yritysvetoinen tekoälyn etiikka kehittyikin suureksi osin avustamaan dataetiikan mukaisissa kysymyksissä.¹¹⁷ Dataetiikka on tärkeä osa-alue tekoälyn etiikassa, mutta sen ottaminen etiikan lähtökohdaksi on osaoptimointia. Osaoptimointi on nähtävissä kysymyksen asettelussa. Dataetiikka vastaa enemmän kysymykseen, kuinka tekoälyä tulisi hyödyntää kuin mitä hyödyntämisellä tulisi tavoitella. Dataetiikka ei problematisoi tekoälyn hyödyntämisen kohteita, vaan tapoja. Näin ollen se ei aseta päämääriä, vaan keinoja.¹¹⁸

Laitila on samaa mieltä Ollilan kanssa todetessaan, että globalisaatio on muuttanut maailman markkinavetoiseksi, mikä on puolestaan vinouttanut ihmisten arvomaailmaa. Hararin mukaan yksi vaikuttava tekijä vinoumien muodostumiselle on ollut, että talousjärjestelmän kehityksellä oli näkyvä rooli ihmisten hyvinvoinnin lisääntymisessä 1800- ja 1900-luvuilla. Teollisten vallankumousten myötä yhteiskuntien taloudellinen epätasa-arvo

¹¹⁴ Harari 2017, 27, 242; Laitila 2019, 16; Ollila 2019, 98.

¹¹⁵ Schauer 2015; Weeden & Brusky 2013; Tieteen termipankki 2015.

¹¹⁶ Ollila 2019, 108.

¹¹⁷ Ollila 2019, 98, 128.

kaventui ja yhä suurempi osa ihmiskuntaa pääsi pois nälänhädän ja epäpuhtaiden elinolojen aiheuttamien sairauksien vaikutusten piiristä. Lopulta taloudellinen kehitys yhdistettiin ihmisten hyvinvointiin niin vahvasti, että bruttokansantuotteesta (bkt) tuli kansallisen hyvinvoinnin mittari. Vielä tänäkin päivänä bruttokansan tuotetta (bkt) käytetään yhtenä kansantalouden hyvinvoinnin ja onnellisuuden mittarina. Yksilön kokemusta ja vapautta korostava liberalismi vaikutuksesta taloudellisesta varallisuudesta on kuitenkin kehkeytynyt itseisarvo, joka rinnastetaan hyvinvointiin.¹¹⁹

Chicagon yliopiston oikeustieteellisen laitoksen etiikan professori Martha Nussbaum argumentoi teoksessaan *Taloukskasvua arvokkaampaa* (2014), että kansalaisten koulutuksen monipuolisuus on keskeinen tekijä yhteiskunnalliselle hyvinvoinnille ja koheesiolle. Hänen mukaansa taloukskasvun tavoittelu on aiheuttanut koulutusjärjestelmien tavoitteiden vinoutumista Yhdysvalloissa, jossa koulutuksen aiempi päämäärä on ollut laaja-alainen sivistäminen.¹²⁰ Nyt sen tavoitteet ovat kaventumassa palvelemaan matemaattis-luonnontieteiden, teknillisten tieteiden ja taloustieteiden tarpeita. Kaventuminen aiheutuu Nussbaumin mukaan pitkälti siitä, että kyseisillä aloilla nähdään olevan selvimät kausaalisuhteet Yhdysvaltojen kansantalouden kasvun kanssa.¹²¹

Koulutuksen nykyinen suunta vaarantaa demokratian perusarvojen, kuten myötätunnon, empatian ja altruismin mukaisen ajatusmaailman kehittymistä, joita humanististen ja taideaineiden opetus tukevat. Kapea päämäärä tuottaa kapeaa ajattelua, ja empatia ei kukoista kapeissa todellisuuskäsityksissä. Empatia on Nussbaumin mukaan keskeinen tekijä elämän itseisarvon ymmärtämisessä.¹²²

Ollilan kuvauksen mukaan markkinoiden sisällyttäminen päätöksentekoon aiheuttaa aina eettisen ajattelun kapea-alaistumista. Tämän vuoksi hän välttäisi antamasta markkinatoimijoille, kuten yrityksille vastuuta tekoälyn etiikan kehittämisessä.¹²³ Todellisuudessa talousjärjestelmä on kuitenkin myös mahdollistanut paljon hyviä asioita, kuten Hararin esimerkit köyhyyden ja nälänhädän vähentymisestä osoittavat.

¹¹⁸ Dataetiikan periaatteista on johdettavissa sen taustalla vaikuttavia arvoja, mutta toimiessaan keskustelun lähtökohtana dataetiikka kaventaa eettisen pohdinnan kattamaan vain pientä osaa tekoälyn etiikasta. Ollila 2019, 97.

¹¹⁹ Harari 2017, 27, 39; Laitila 2019, 16.

¹²⁰ Nussbaum tiedostaa, että Yhdysvalloissa on paljon haasteita koulutuksen epätasaisuuden kanssa. Hän tarkastelee kirjassaan koulutusjärjestelmän suuria linjoja.

¹²¹ Nussbaum 2011, 15.

¹²² Nussbaum 2011, 15, 43, 58.

¹²³ Ollila 2019, 94–98.

Talousjärjestelmä on mahdollista ymmärtää hyödykkeiden ja palveluiden koordinoitijärjestelmänä. Sellaisena se on välttämätön osa yhteiskuntien kokonaisuutta.

Laitila ja Harari muistuttavat, että juuri vapaiden markkinoiden kehityksellä on ollut myös ihmisiä yhtenäistävää vaikutusta globaalilla tasolla. Se on ollut muuttamassa ihmisten ajatusmaailmaa yhteistyöstä niin kutsutuksi yhteisvoiton (win-win) tilanteeksi, aiemmin vallinneesta ajatuksesta, jonka mukaan yhteistyössä on aina voittava ja häviävä osapuoli.¹²⁴ Nussbaumin, Laitilan ja Hararin ajatus on, että talousjärjestelmä tulisi ymmärtää osana yhteiskunnan ekosysteemiä, mutta ei päämääränä itsessään. Tämä pätee puhuttaessa niin tekoälystä kuin mistä tahansa muustakin yhteiskunnallisesta kehityksestä.

¹²⁴ Harari 2017, 27, 379; Laitila 2019, 71.

3 Antroposeenin etiikka

Sekä Maija-Riitta Ollila että Erkki Laitila toteavat, että tekoälyn kehittämisen tavoitteiden ja toimintatapojen valinnassa ei tule jäädä kansalliselle tasolle. Laitilan perustelu on, että informaatioteknologian ja vapaan globaalin talousjärjestelmän kehittymisen myötä maailma muodostaa suuren informaatio-, valuutta- ja kulttuuriverkoston, jossa yksikään valtio ei ole muista eristyksissä. Näin ollen vain kansallisen tason tekoälykehitykseen keskittyminen olisi osaoptimointia. Hän ehdottaa tekoälyn kehittämisen globaaleiksi tavoitteiksi megatrendien muodostamien haasteiden ratkaisemista. Laitilan ehdottamat megatrendit ovat peräisin Osmo Kuusen ja Risto Linturin tulevaisuuden innovaatioita esittävästä tutkimuksesta. Keskeisimmät Linturin ja Kuusen mainitsemista megatrendeistä ovat ilmastonmuutos, globalisaatio, eriarvoisuuden lisääntyminen, arvomaailman vääristynyt kehitys ja tieteen ja tekniikan yhdentyminen.¹²⁵

Ollilan näkemys kansallisen tason etiikan heikkoudesta perustuu sille, että etiikkaan kuuluu olennaisesti ajatus inklusiivisuudesta. Jos eettisen pohdinnan lähtökohdaksi otetaan ajatus, että jokin ryhmä on parempi tai tärkeämpi kuin toiset, päädytään sisäisesti ristiriitaisiin toimintaperiaatteisiin. Lisäksi tekoälyn etiikka koskee teknologiaa, jonka vaikutukset ovat maailmanlaajuisia, joten eettisen pohdinnankin tulisi ottaa globaali konteksti huomioon. Haasteeksi muodostuu löytää kaikkia toimijoita yhdistävä arvopohja, josta voitaisiin systemaattisesti johtaa toimintaperiaatteita. Juuri tämä vaikeus on yksi syy, miksi tekoälyn etiikka muotoillaan usein keskitason periaatteiksi. Ollila argumentoi, että eettinen pohdinta edellyttää selvemmän teoreettisen pohjan kuin liberaalien arvojen kompromisseja ilmentävät keskitason periaatteet tarjoavat.¹²⁶ Yksi ihmisiä, eläimiä ja ympäristöä yhdistävä tekijä on kuitenkin helppo nimetä: elämä.¹²⁷

Ollila ja Harari jäsentävät ihmisen ja luonnon suhdetta antroposeenin käsitteen avulla. Antroposeeni tarkoittaa geologista aikakautta, jolloin ihminen on suurin yksittäinen vaikuttaja maapallon ympäristön muokkautumiseen.¹²⁸ Ajatus kuulostaa äkkiseltään erikoiselta ja geologit ovat erimielisiä siitä, onko maailma tosiasiassa siirtynyt edellisen jääkauden jälkeen alkamaan ajoitetusta holoseenista antroposeeniin.¹²⁹ Hararin ja Ollilan perustelut siitä, että

¹²⁵ Laitila 2019, 15; Linturi ja Kuusi 2018, 400.

¹²⁶ Näistä huomioista huolimatta Ollila arvostaa muun muassa Future of Life Instituten Asilomar -periaatteita ja Montrealin yliopistossa kehitettyä *Montreal Declaration for Responsible AI* -julistusta.

¹²⁷ Ollila 2019, 24, 46, 99.

¹²⁸ Waters et al. 2016; Harari 2017, 80; Ollila 2019, 21.

¹²⁹ Waters et al. 2016.

elämme antroposeenissa täydentävät toisiaan. Ollila tarkoittaa antroposeenilla ihmisten vaikutuksia teollisten vallankumousten jälkeisellä aikakaudella. Hän käyttää käsitettä erityisesti kuvatessaan ihmisten aiheuttamia ilmastopäästöjä ja niiden välittömiä ja välillisiä globaaleja vaikutuksia.¹³⁰

Harari puolestaan kuvaa antroposeenilla laajempaa toimintatapojen kattausta, jolla ihminen on vaikuttanut asuttamiinsa ekosysteemeihin. Harari tarkastelee ihmisen toimintaa 70 000 vuoden jaksolta, ja kuvaa, kuinka jo ensimmäiset homo sapiensit vaikuttivat toimintaympäristöihinsä tuhoisin seurauksin. Ihminen on lajinsa alkuaikoina vaikuttanut 70 muun ihmislajin katoamiseen ja myöhemmin usean muun eläinlajin sukupuuttoon kuolemiseen, jo ennen siirtymistä agraariyhteiskuntiin ja sitä kautta kehittyneemmän teknologian aikaan.¹³¹

Vaikka ihminen vaikutti asuttamiinsa ekosysteemeihin suuresti jo tuhansia vuosia sitten, ihminen ei lajina ymmärtänyt toimintansa seurauksia. Vielä teollisten vallankumoustenkaan aikaan ihmiset eivät hahmottaneet, että he olivat keskeinen tekijä ekosysteemitasoisissa muutoksissa. Logistiikkaketjujen laajentuessa kattamaan suurimman osan maailmasta, ihminen vaikutti useiden tappavien epidemioiden leviämiseen mantereelta toiselle. Antroposeenin myötä maailmasta on tullut ensimmäisen kerran yksi ekologinen yksikkö.¹³²

Tärkeä käännekohta Hararin rakentamassa narratiivissa tapahtui 1900-luvun aikana. Tuolloin ihminen tuli tietoiseksi tahattomista vaikutuksistaan ympäristöönsä ja myös mahdollisuuksistaan vaikuttaa kehitykseen teknologian avulla. Nykyään inhimillisen toiminnan ekosysteemi kattaa koko maailman. Talouden, logistiikan ja viestinnän globalisoitumisen myötä ei ole paikkoja, joissa ihminen ei vaikuttaisi, eikä vaikutus ole läheskään aina positiivinen. Ihminen on vaikuttanut muun muassa ilmastomuutoksen kiihtymiseen, maailman metsäpeitteen katoamiseen, merten lämpenemiseen, eläinkantojen katoamiseen ja valtasuhteiden muutokseen eläinkannoissa. Valtasuhteilla Harari tarkoittaa sitä, että 2010-luvulla villieläimet kattavat enää alle 10% kaikista maailman suureläimistä¹³³. Loput yli 90% koostuu ihmisistä ja domestikoiduista eläimistä.¹³⁴ Kansainvälisen ilmastopaneelin IPCC:n vuonna 2018 julkaisema tutkimusraportti tukee Ollilan ja Hararin

¹³⁰ Ollila 2019, 22.

¹³¹ Harari 2017, 82.

¹³² Harari 2017, 82.

¹³³ Suureläimellä tarkoitetaan yli kaksi kiloa painavia eläimiä.

¹³⁴ Harari 2017, 80.

ajatuksia antroposeeniin siirtymisestä. Raportissa todetaan, että inhimillisen toiminnan vaikutus ilmastomuutoksen kiihtymiseen ja eliöiden sukupuuttoihin on kiistatonta.¹³⁵

Harari kuvaa antroposeenia historian tutkimuksen tapaan deskriptiivisesti. Ollila puolestaan näkee antroposeenin velvoittavan ihmistä. Kun ihmisen toiminnalla on selkeä vaikutus maapallolla olevan elämän jatkuvuuteen, on ihmisellä myös velvollisuus pyrkiä positiiviseen vaikutukseen toiminnallaan. Ollilan mukaan elämän jatkuvuuden varmistaminen on näin ollen antroposeenin etiikan lähtökohta. Myös Laitilan mukaan ilmastomuutoksen ratkaisemisen tulisi olla tekoälyn kehityksen keskeisimpiä tavoitteita. Tietotekniikan päätavoite on auttaa ratkaisemaan ihmisten ongelmia, ja ilmastomuutos on eräs haastavimmista ongelmista, joita ihmiskunta on kohdannut.¹³⁶

3.1 Yhteistä toimintaa rakentamassa

Antroposeenin etiikka asettaa kestävän kehityksen eettisen pohdinnan keskeisimmäksi arvoksi. Arvot tarvitsevat jatkojalostusta periaatteiksi ja toimintaohjeiksi, jotta ne voisivat aktualisoitua inhimillisessä toiminnassa.¹³⁷ Tarkastelen seuraavaksi yritys vastuuteorioille pohjautuvaa ajattelua inhimillisen toiminnan ohjaajana. Maija-Riitta Ollila pohtii teoksessaan, miltä näyttäisi yritysveltoinen tekoälyn etiikka. Hän hyödyntää pohdinnassaan Immanuel Kantin ajatuskoetta arvojen valtakunnasta, jossa moraalisen toiminnan subjektit ovat tasaveroisessa asemassa ja päättävät yhdessä moraalisen toiminnan määrittävät arvot. Ollila tarkastelee arvojen valtakuntaa yhteiskuntavastuun käsitteen avulla. Yhteiskuntavastuulla tarkoitetaan yritysten osallistumista yhteiskunnan sosiaalisten ja ekologisten ongelmien ratkomiseen.¹³⁸ Ollila argumentoi Hararin ja Laitilan tapaan, että yritysveltoinen arvojen määrittely päättyy todennäköisimmin osaoptimointiin, jossa taloudellisen hyödyn tavoittelu on varsinainen päämäärä ja muut arvot määrittyvät hierarkkiseen järjestykseen sen mukaan, kuinka ne palvelevat varallisuuden karttumista. Yritykset hyödyntävät yhteiskuntavastuuta sekä viherpesuna että välttääkseen valtion hallintojen säätämiä yritystoimintaa koskevia lakeja.¹³⁹

Yritysten edustajien tekemät tekoälyn eettiset periaatteet ovat Ollilan mukaan hyviä esimerkkejä yhteiskuntavastuun tapaisesta viherpesusta. Ne muotoillaan usein keskittymään dataetiikkaan ja vaikka dataetiikka on tärkeä aihe tekoälyn etiikassa, niin se edustaa vain osaa

¹³⁵ IPCC 2018; Ollila 2019, 22.

¹³⁶ Harari 2017, 77–78; Laitila 2019, 49.

¹³⁷ Porter & Kramer 2011; Ollila 2019, 107.

¹³⁸ Ollila 2019, 95.

¹³⁹ Ollila 2019, 94–98.

siitä. Muun muassa Microsoft ja Google ovat lanseeranneet omat periaatteensa ja Facebook kuuluu yritys yhteistyölle perustuvaan eettistä tekoälyä tutkivaan *Partnership on AI* yhteisöön.¹⁴⁰ Näistä kolmesta monikansallisesta yrityksestä Facebook ja Google ovat ajautuneet hankaluuksiin vuodettuaan käyttäjätietojansa 2010-luvulla. Vuonna 2018 kummankin yrityksen toiminnasta paljastui suuria tietovuotoja, joita päättävässä asemassa olevat yrityksen työntekijät ja osakkaat olivat pyrkineet peittelemään.¹⁴¹

Vuoden 2018 tammikuussa paljastuneessa tietovuodossa data-analyysiin erikoistunut yhtiö Cambridge Analytica pääsi käsiksi miljoonien ihmisten Facebook-tilitietoihin, joita se sittemmin hyödynsi poliittisen mainonnan kohdentamiseen Yhdysvaltain vuoden 2016 presidentinvaaleissa.¹⁴² Myös Googlen tietovuoto johti miljoonien ihmisten henkilökohtaisten tietojen päätymiseen kolmannen osapuolen käytettäväksi ilman käyttäjien hyväksyntää. Facebookin ja Googlen johtavat tahot olivat tietoisia vuodoista ja pyrkivät niiden ehkäisemisen sijaan peittelemään jälkiään.¹⁴³ Toiminta on hyvinkin ristiriitaista Googlen ja *Partnership on AI* yhteisön eettisten periaatteiden kanssa, joissa datan turvallisuus ja yksilön suoja ovat merkittävässä roolissa.

Ollilan kuvaama yhteiskuntavastuu asettaa sosiaalisen hyvän tavoittelun yritysten sivutoiminnaksi, jolla pyritään välttämään valtion hallinnon puuttumista yritystoimintaan. Ollilan ajatus, että tekoälyn etiikkaa tulisi kehittää pitkälti yritysmaailman ulkopuolella, on kuitenkin osaoptimointia, sillä yritykset ovat keskeinen taho tekoälyn kehittäjinä. Etiikan siirtyminen käytäntöön olisi hyvin hidasta ja epävarmaa, ellei yritystoimintaa osallisteta suoraan tekoälyn etiikan keskusteluihin, vaan niiden toimintaan pyrittäisiin vaikuttamaan välillisesti ohjelmoinnin etiikan kehittämisen kautta, kuten Ollila ehdottaa.¹⁴⁴

Mitä jos eettinen pohdinta tuottaisikin yrityksille mahdollisuuden hyödyntää ydinosaa yhteiskunnan ja yrityksen hyväksi? *Creating Shared Value (CSV)* -teoria muuttaa sosiaalisen ja ekologisen hyvinvoinnin tavoittelun yhteiskuntavastuusta yrityksen kilpailueduksi. Ensimmäisinä ajatuksen CSV -teoriasta esittivät Harvardin yliopiston liiketaloustieteen professori Michael E. Porter ja FSG:n¹⁴⁵ toimitusjohtaja Mark Kramer

¹⁴⁰ Microsoft ja Google kuuluvat myös *Partnership on AI*:hin. Facebook ja Microsoft kuuluvat sen perustajayrityksiin. Novet 2018; Pichai 2018; Facebook s.a.; Microsoft AI Principles s.a..

¹⁴¹ Räisänen 2018; Hern 2018; Hern 2019.

¹⁴² Tietovuoto ilmeni suurelle yleisölle vuoden 2018 tammikuussa ja sen seurauksena Cambridge Analytica ajautui konkurssiin.

¹⁴³ Räisänen 2018; Hern 2018; Hern 2019; Ollila 2019, 96.

¹⁴⁴ Ollila 2019, 112.

¹⁴⁵ FSG on konsulttiyritys, joka keskittyy auttamaan organisaatioita löytämään tapoja parantaa yhteiskunnan sosiaalista hyvinvointia toiminnallaan.

Harvard Business Reviewn artikkelissaan *Creating Shared Value* (2011). Heidän mukaansa CSV ei ole vain uusi kuvaus yritysten yhteiskuntavastuusta, vaan se edustaa kapitalismin systeemitason muutosta. Sen pyrkimys on saada ihmiset hahmottamaan talousjärjestelmä yhteiskunnallisten tarpeiden tyydyttämisen välineenä.¹⁴⁶ Näkemyksellä on yhtäläisyyksiä Hararin ja Laitilan talousjärjestelmän käsitteen kuvausten kanssa.

Kapitalistisen talouskehityksen aiheuttamat markkinahäiriöt ovat johtaneet yritystoiminnan legitimitietin kyseenalaistamiseen yhteiskunnallisissa keskusteluissa. Tilanteen korjaantumiseksi yritystoimintaa harjoittavien tulisi ensisijaisesti tähdätä tuottamaan hyvinvointia koko ekosysteemin tasolla, jonka onnistuminen avaa yritykselle uusia markkina-alueita ja vahvistaa sen kilpailuetua. Porter ja Kramer tähdentävät, että CSV ei tarkoita, että yritysten roolin pitäisi muuttua julkishallinnon kaltaiseksi yhteiskunnallisten ongelmien rahoittajaksi ja ratkaisijaksi. Yritystoiminnan tulisi kuitenkin perustua laajempaan yhteistyöhön, jonka tavoite on korjata markkinahäiriöitä. *Anglo American* -kaivosyhtiön alulle saattama projekti Etelä-Afrikan AIDS epidemian taltuttamiseksi, monikansallisen ravinneyrityksen *Yaran* aloite, jolla parannettiin ruoan tuotantoa sekä logistiikkaa nälänhädästä kärsivässä Tansaniassa, ja *Walmartin* aloitte, joka paransi Yhdysvaltojen jätteenlajitteluinfrastruktuuria vähentäen vuosittaisen muovijätteen määrää 800 000 tonnilla (kg) ja kasvihuonekaasujen määrää 200 000 tonnilla (kg), ovat käytännön esimerkkejä CSV:n mukaisesta toiminnasta. Kaikissa tapauksissa yritykset olivat aloitteen tekijöitä kansallisen tason hyvinvointia edistävissä projekteissa ja kasvattivat samalla omaa markkina-arvoaan.¹⁴⁷

Anglo Americanin toiminnassa yhteiskunnallisen hyvän ja yrityksen markkina-arvon kohennuksen yhteys on helppo havaita, sillä panostamalla epidemian taltuttamiseen, yhtiö varmisti, että sen työntekijät kykenivät tekemään töitä. *Yaran* ja *Walmartin* tapauksissa yhteiskunnalle tuotetun hyvän ja yrityksen markkina-arvon kohentumisen korrelaationsuhde oli mutkikkaampi. *Walmartin* johtajat halusivat korvata yrityksen tuottamaa muovijätteen määrää käyttämällä jätemuovia pakkausmateriaaliensa ja muovipussiensa valmistukseen. Ongelmaksi muodostui, että Yhdysvalloissa ei ollut kunnollista muovinkierrätysjärjestelmää, minkä vuoksi raaka-ainetta ei ollut saatavilla. Lanseeraamansa projektin avulla Walmart sai halpaa materiaalia muovin valmistukseen ja samalla kansallisen tason ympäristöhaitat pienenevät.¹⁴⁸

Yaralla oli puolestaan vaikeuksia saavuttaa Afrikan pienviljelijöiden markkinoita Etelä-Tansanian satama-aluetta vaivaavan korruption, maan huonon tieinfrastruktuurin,

¹⁴⁶ Porter & Kramer 2011, 4.

¹⁴⁷ Porter & Kramer 2011, 4–5; Kramer & Pfitzer 2016.

kuljetusvälineiden puutteen ja maanviljelijöiden köyhyyden sekä ravinnevastamielisyyden vuoksi. Yhtiön johto kuitenkin uskoi, että sen tuottamalla ravinteilla olisi Tansaniassa kysyntää maan, ja sen lähialueiden köyhän viljelysmaaperän vuoksi. Niinpä Yara perusti *Southern Agricultural Growth Corridor of Tanzania* (SAGCOT) -hankkeen. Yhtiö kutsui koolle asiantuntijoita 68:sta organisaatiosta, joiden joukkoon kuului kehitysapujärjestöjen ja kansalaisoikeusjärjestöjen työntekijöitä, Tansanian hallituksen ministereitä sekä muiden monikansallisten yritysten johtoa. SAGCOT:n tavoitteeksi asetettiin kehittää maanviljelysjärjestelmä täysin uudelleen aina Etelä-Tansanian satamasta, maan läntiselle rajalle asti. Alue on pinta-alaltaan noin Italian suuruinen.¹⁴⁹

Yara fasilitoi SAGCOT:n alullepanon, mutta sulautui sen jälkeen muiden toimijoiden sekaan. Hanke lähti hyvin käyntiin ja sille nimettiin autonominen toimintaa koordinoiva organisaatio, joka sai rahoitusta hankkeeseen osallistuvilta yrityksiltä sekä Tansanian valtiolta. SAGCOT -projektin arvioitiin kustantavan noin 3.4 miljardia euroa. Julkinen rahoitus kattoi siitä kolmasosan ja loput kulut jaettiin osallistuvien yritysten kesken. Kolme vuotta SAGCOT:n perustamisen jälkeen satama-alue ja tarvittavat tie- ja sähköinfrastruktuurit oli uudistettu, uusi ravinnejalostamo rakennettu, korruption vaikutus satama-alueella vähentynyt ja satojen tuhansien maanviljelijöiden taloustilanne oli parantunut. Loppujen lopuksi Yara rahoitti projektia 60 miljoonan euron verran parantaen samalla tuottavuuttaan SAGCOT:n vaikutusalueella 50%.¹⁵⁰

CSV ohjaa yrityksiä hahmottamaan yhteiskunnallista hyvää ja etsimään tapoja, joilla voisi tukea sen saavuttamista. Porter ja Kramer hyödynsivät *kollektiivisen vaikuttamisen* (Collective impact) teoriaa CSV:n yhtenä kulmakivenä. John Kania ja Mark Kramer hahmottelivat kollektiivisen vaikuttamisen teorian tutkiessaan laajamittaiseen sosiaaliseen muutokseen tähdänneitä kansalaisjärjestöjen projekteja. Teorian mukaan onnistunut sosiaalinen muutos tapahtuu vain, jos yhteisön sisäiset ryhmät tekevät sektorit läpileikkaavaa systemaattista yhteistyötä.¹⁵¹

Kania ja Kramer argumentoivat, että nykyiset kansalaisjärjestöjen rahoitusmekanismit luovat järjestöjen välille kilpailuasetelman yhteistyön pohjustamisen sijaan. Samalla rahoittajat jäävät ulkopuolisen tukijan asemaan, jolloin sijoitukselta toivotaan mahdollisimman skaalautuvaa ratkaisua. Toisin sanoen rahoitusmalli sisältää oletuksen, että

¹⁴⁸ Kramer & Pfitzer 2016, 4–8.

¹⁴⁹ Kramer & Pfitzer 2016, 4–8.

¹⁵⁰ Kramer & Pfitzer 2016, 6–8.

¹⁵¹ Kania & Kramer 2011.

yksittäinen organisaatio kykenisi ratkaisemaan kompleksisia ja yhteiskunnallisia sektoreita läpileikkaavia ongelmia. Teorian mukaan laadukas yhteistyö sisältää viisi vaatimusta: sillä tulee olla yhteinen tavoite, yhteiset mittarit, toisiaan täydentävää toimintaa, jatkuvaa kommunikaatiota ryhmien kesken ja projektille nimetty operoiva organisaatio. Kollektiivisen vaikuttamisen mukaista toimintaa harjoittaneet projektit ajautuivat Kanian ja Kramerin mukaan taloudellisiin vaikeuksiin hyvistä saavutuksistaan huolimatta. Porterin ja Kramerin mukaan CSV:n tapa sisällyttää yritystoiminta osaksi yhteisen hyvän toteuttamista korjaa hankkeiden rahoitusvaikeuksia. Yritysten mukanaolo tuo hankkeisiin rahoituksen lisäksi tärkeää osaamista toiminnan koordinoimiseksi ja skaalautuvuuden varmistamiseksi.¹⁵²

CSV hyödyntää holarkian mukaista ekosysteemin ajatusta yritystoiminnan toimintaympäristön hahmottamiseksi. Mark Kramer ja Mark Pfitzer kehittivät edelleen CSV:n käytännön toimeenpanon teoriaa artikkelissaan *The Ecosystem of Shared Value* (2016). Heidän mukaansa CSV:n ekosysteemin keskeiset toimijat koostuvat muista yrityksistä, valtionhallinnosta, kansalaisjärjestöistä ja kansalaisista.¹⁵³ Mielestäni akateemiset oppilaitokset, eli korkeakoulut olisi hyvä lisätä keskeisten toimijoiden joukkoon. Ensinnäkin oppilaitokset edustavat aluetta, joka yhdistää suurta osaa ihmisistä ennen työelämän sektoreille jakautumista. Koulutuksen kautta voidaan näin ollen vaikuttaa laaja-alaisesti. Toisekseen akateeminen tutkimus pyrkii objektiivisuuteen, yritysten ja järjestöjen tutkimusten lähtökohtien ollessa tiukemmin kytköksissä organisaation ideologiaan. Akateemiset laitokset tuottavat CSV:n mukaisten hankkeiden aiheista syventävää tutkimusta, jota muun muassa ilmastonmuutoksen kokonaisvaltaisten ratkaisujen luominen edellyttää.¹⁵⁴

Kramerin ja Pfitzerin mukaan CSV auttaa etenkin markkinavirheiden ratkaisemisessa, sillä niiden juurisyöt ovat monisyiset ja vaativat kollektiivisen vaikuttamisen kaltaista yhteistyötä.¹⁵⁵ Tekoälyn kehitys on potentiaalisesti johtamassa negatiivisiin ulkoisvaikutuksiin¹⁵⁶, joista työelämän muutos on konkreettinen esimerkki. Tämän lisäksi tekoälyllä on potentiaalia tehostaa kestävän kehityksen mukaista toimintaa, jos sen hyödyntämistä kohdennettaisiin oikein. Näin ollen tekoälykehityksellä on yhteyksiä markkinavirheisiin ja CSV voisi tukea sen eettistä hyödyntämistä.

¹⁵² Kania & Kramer 2011.

¹⁵³ Kramer & Pfitzer 2016, 4–6.

¹⁵⁴ Younglong et al. 2015.

¹⁵⁵ Kramer & Pfitzer 2016, 11.

¹⁵⁶ Ulkoisvaikutuksilla tarkoitetaan ihmisen toiminnasta aiheutuvia sosiaalisia tai ekologisia haittoja, joista kukaan ei ole laillisesti vastuussa.

Porter ja Kramer ovat rakentaneet teoriansa ohjaamaan yritystoiminnan päätöksentekoa. En kuitenkaan näe syytä, miksi muut ekosysteemin keskeiset tahot eivät voisi olla aloitteen tekijöitä. Aloitteen tulee saada tarpeeksi yhteiskunnallista auktoriteettia taakseen, jotta se tulee kuulluksi. Lisäksi aloitteen tekijään kohdistuu vaatimus, että hän osaa argumentoida perusteellisesti, miksi aloitteen tavoite on tarpeellinen, myös kunkin osallisen omasta näkökulmasta tarkasteltuna. Väitän, että mikä tahansa CSV:n sosiaalisen ekosysteemin (kansalaiset, valtion hallinto, yritykset, kansalaisjärjestöt ja korkeakoulut) tahoista voi täyttää edellä mainitut ehdot. CSV voisi näin ollen tarjota valtioiden hallituksille konkreettisen tekoälyn eettisen hyödyntämisen työkalun.

3.2 Aikaraamin vaikutus arvoihin

”Etiikassa tulevaisuus ulottuu vain siihen asti, mihin ajattelu yltää.”¹⁵⁷ Ollila kiteyttää tässä lauseessa hyvin, miksi toiminnan tarkastelun aikaraami vaikuttaa valittuihin päämääriin. Hän toteaa, että seuraavaan kvartaaliin tähtäävällä yhtiön johtajalla on hyvin erilaiset päämäärät toiminnalleen kuin henkilöllä, joka tähtää ajattelussaan Auringon sammumiseen asti. Ajatuksessa näkyy etiikan pitkäkatseisuuden ja inhimillisen toiminnan vaatimusten välinen ristiriita, sillä suhteellisen lyhyet ja selkeät aikaraamit ovat tärkeä osatekijä inhimillisen toiminnan päämäärien saavuttamiseksi. Antroposeenin etiikka asettaa erityisen vaatimuksen aikaraamille. Ihmisen on kyettävä tarkastelemaan toimintansa seurauksia kestävän kehityksen vaatimusten mukaisesti yli sukupolvien.¹⁵⁸

Kuten antroposeenin etiikka, myös CSV edellyttää pitkän aikavälin tavoitteet. Tavoitteiden aikajänteessä konkretisoituu eettisen kehityksen ja vapaan markkinatalouden välinen ristiriita. Yksi keskeinen haaste toteuttaa CSV:n mukaista yritystoimintaa on, että sillä ei voi luvata varmoja lyhyen tähtäimen voittoja eikä kuuluisikaan voida. Kyseinen piirre saattaa kuitenkin näyttäytyä osakkeen omistajille suurena riskinä. Jos yritysjohtajat tekevät päätöksiä sen mukaan, mitä ajattelevat osakeomistajien toivovan, he ajautuvat ajattelemaan kvartaalien mukaan. Porterin ja Kramerin mukaan CSV:n mukainen kehitys on lopulta toimintaympäristön hyvinvoinnin takaamisen lisäksi myös kapitalismin mukaista voitontavoittelua parempi tapa varmistaa vaurautta osakkeenomistajille. SAGCOT hanke on tästä hyvä esimerkki. Taloudellinen tuottavuus on CSV:n sivutuote. Teoria ohjaa ajattelemaan, että kilpailuedut eivät olekaan enää yrityskohtaisia, vaan ekosysteemikohtaisia ja yritykset eivät välttämättä ole kilpailijoita keskenään vaan osa toistensa ongelmien

¹⁵⁷ Ollila 2019, 12.

ratkaisua.¹⁵⁹ CSV tukee siltä osin Laitilan ja Hararin mainitsemaa yhteisvoiton mukaista ajattelua tavanomaista vapaata markkinatalouden mukaista kehitystä vahvemmin.

Tekoälyn etiikan keskustelu on pitkälti ohjautunut keskinäisten periaatteiden mukaiseksi ja keskittynyt ohjelmoinnin- ja dataetiikkaan. Antroposeenin etiikka ja CSV:n mukainen työskentely voisivat mahdollistaa yhteiskunnallisen arvopohjan ja päämäärien asettamisen, joista voitaisiin puolestaan johtaa toimintaa ohjaavia periaatteita ja käytäntöjä. Tarkastelen seuraavassa osiossa neljän valtion hallitusten laatimista tekoälystrategioista heijastuvaa eettistä pohdintaa. Viitekehyksenä toimii tähän asti tarkastellut holarkian, ja affordanssin käsitteet, CSV ja antroposeenin etiikka.

¹⁵⁸ Ayres et al. 1998, 1; IPCC 2018; Ollila 2019, 12, 28.

¹⁵⁹ Porter & Kramer 2011, 4–6, 17; Kramer & Pfitzer 2016, 10.

4 Strategioiden ilmentämä etiikka

Tapa, jolla tekoälyn kehityksestä puhutaan vertailtavina olevissa tekoälystrategioissa, antaa hyvän kuvan, minkälaisista lähtökohdista strategioita on lähdetty rakentamaan. Ison-Britannian hallituksen työryhmän laatima raportti *Artificial Intelligence Sector Deal*¹⁶⁰ (2018) selittää tekoälyn kehityksen positiivisessa valossa ja välttää kuvaamasta tekoälykehityksen sisältämiä riskejä. Riskejä ei mainita, vaan niihin viitataan puhumalla dataetiikan tarpeellisuudesta. Ainoa riski, joka ilmaistaan suoraan, on aivovuodon mahdollisuus, mikäli hallitus ei ottaisi tekoälystrategiaa käyttöön. Aivovuodon riski mainitaan Yhdysvaltojen hallituksen työryhmän raportissa *Summary of the 2018 White House summit on Artificial Intelligence for American Industry*¹⁶¹ (2018). Työryhmä ei niinkään puhu tarpeesta ottaa tekoälystrategia käyttöön vaan tarpeesta olla säännöstelemättä tekoälyn kehitystä liikaa, jotta yritysten innovaatiomahdollisuudet säilyvät.¹⁶²

Ranskan hallituksen työryhmän *Villani Report*¹⁶³ (2018) ja Suomen hallituksen työryhmän *Tekoällyajan työ – Neljä näkökulmaa talouteen, työllisyyteen, osaamiseen ja etiikkaan*¹⁶⁴ (2018) -raporteissakin mainitaan aivovuodon riski, mutta se ei ole yhtä keskeinen tekijä kuin AI Sector Deal ja White House summit on AI -raporteissa. Villani ja Tekoällyajan työ -raporteissa puhutaan sen sijaan tarpeesta luoda selkeä visio tekoälyn hyödyntämisestä, minkä avulla valtio voi ohjata tekoälyn kehitystä toivottuun suuntaan. Visioiden tulee pohjautua eettisesti kestäväälle pohjalle, mikä tarkoittaa sekä tekoälyn hyödyntämisen sisältämien riskien että käyttömahdollisuuksien tunnistamista. Raporteissa esitettyjä riskejä ovat muun muassa työttömyys, epätasa-arvon voimistuminen, datavääristymät, tekoälyn väärinkäyttö, globaalin poliittisen vallan siirtyminen suuryrityksille ja vallanhimoisen tai vahinkoa tuottavan vahvan tekoälyn kehittyminen.¹⁶⁵

Tekoälyn potentiaali tuottaa hyvää tunnistetaan kaikissa raporteissa, mutta siitä puhutaan esimerkiksi White House Summit on AI -raportissa lähes yksinomaan yhteiskunnallisten sektorien toiminnan tehostamisen näkökulmasta. Hyvän rinnastaminen tehostamiseen kuvastaa hyvin Cicmillin ja Hodgsonin mainitsemaa arvovinoumaa, jossa toiminnan tehostaminen on muuntunut itseisarvoksi. Toisaalta raportin kirjoittajien mukaan

¹⁶⁰ Tästä edespäin AI Sector Deal.

¹⁶¹ Tästä edespäin White House Summit on AI.

¹⁶² Clark et al. 2018, 3, 9; OSTP 2018, 11.

¹⁶³ Tästä edespäin Villanin raportti.

¹⁶⁴ Tästä edespäin Tekoällyajan työ.

¹⁶⁵ Koski & Husso 2018, 18, 45; Villani et al. 2018, 6–7, 10, 80, 112.

toiminnan tehostaminen vaikuttaa sektorien taloudelliseen tuottavuuteen, mikä viittaisi Ollilan, Nussbaumin ja Laitilan mainitsemaan arvovinoumaan, jossa taloudellinen tuotto on asetettu itseisarvon asemaan.¹⁶⁶ Tehokkuus ja tuottavuus ovat kumpikin selkeitä välinearvoja eivätkä moraaliarvoja. Muissa raporteissa tekoälyn potentiaali yhdistetään kestävä kehityksen mukaisesti muun muassa ilmastonmuutokseen ja rakennetyöttömyyden haasteisiin vastaamiseen ja mahdollisuuden parantaa terveys- ja hyvinvointipalveluiden saavutettavuutta. Lisäksi Tekoälyajan työ -raportissa mainitaan tekoälysovellusten mahdollisuus parantaa ihmisten kokonaisvaltaista hyvinvointia parantamalla palveluiden laatua ja laskemalla niiden saatavuuden hintaa.¹⁶⁷

Riskien ja kehityksen potentiaalien tunnistamisen perusteella Ison-Britannian, Ranskan ja Suomen hallitusten työryhmät ovat hahmottaneet tekoälyn kehityksen sosio-tekni- senä kokonaisuutena, ja Yhdysvaltojen hallituksen työryhmä on ajatellut yhteiskuntaa ensisijaisesti tekoälykehityksen objektina. Lisäksi White House Summit on AI -raportin perusteella Yhdysvaltain hallituksen rooli on enemmänkin olla kehityksen etäinen mahdollistaja kuin aktiivinen vaikuttaja. Ranskan ja Suomen hallitusten työryhmät ajattelevat CSV:n tapaisesti yhteisen arvopohjan hahmottamisen tärkeäksi lähtökohdaksi, jonka muodostumisessa hallituksella on keskeinen rooli. Asetelma kuvastaa, että raporttien laatijat ymmärtävät tekoälyn kehityksen ja hyödyntämisen yhteiskunnalliset vaikutukset ja valtion roolin niiden toteutumisessa eri tavoin.

4.1 Kansalaisten hyväksyntä minimiarvona

Kaikissa vertailtavissa strategioissa sanotaan, että tekoälykehityksen tulisi tavoitella kansalaisten hyväksyntää. Se, että kansalaiset hyväksyvät valitun suunnan on välttämätöntä, jotta tekoälyä voidaan kehittää ja hyödyntää, mikä on raporttien mukaan edellytys tuottavuuden kasvulle. Kansalaisten hyväksynnän saavuttamisen todellinen tavoite voi siis olla tuottavuuden kasvu. Strategiaraporttien kirjoittajien taustaoletukset saadaan näkyviin, kun tarkastellaan, mistä kansalaisten hyväksynnän ajatellaan koostuvan.

Tekoälyajan työ -raportissa sanotaan, että kansalaisten hyväksynnän saaminen edellyttää, että tekoälyä kehitetään ja hyödynnetään niin, että kansalaisia kohdellaan oikeudenmukaisesti ja kohtuullisesti. Raportin tekijät tarkoittavat oikeudenmukaisuudella sitä, että tekoälyä hyödyntämällä on pyrittävä toteuttamaan koko yhteiskunnan kehitystä eikä vain keskittyttävä tukemaan menestyjiä. Tämä tarkoittaa sitä, että muun muassa

¹⁶⁶ OSTP 2018, 3.

innovaatiopolitiikan tavoitteeksi tulisi asettaa uusien työmuotojen kehittäminen eikä keskittyä työelämän automatisoinnin kiihdyttämiseen. Näin toimimalla voitaisiin lieventää tuloerojen ja työelämästä pois putoamisen aiheuttamaa yhteiskunnallista eriarvoistumista. Tekoälyn innovaatiopolitiikkaa tarkasteltaisiin erityisesti työelämässä heikossa asemassa olevien tukemisena. Kohtuullisuudella tarkoitetaan, että tekoälyteknologian kehitys ei saisi määrittää kansalaisten sopeutumisnopeuden tarvetta, vaan päinvastoin, hyödyntämistään tulisi ottaa kansalaisten sopeutumistarpeet huomioon.¹⁶⁸

Villanin -raportissa todetaan puolestaan, että kehityksen ja hyödyntämisen tulisi pohjautua parhaisiin standardeihin, jotta ne varmasti saavuttaisivat kansalaisten hyväksynnän. Raportin kirjoittajat sanovat tarkoittavansa parhailla standardeilla, että tekoälyn hyödyntämistä varten täytyy kehittää tekoälyn aikakauteen sopivat lailliset ja eettiset raamit. Raamien fokus olisi vastata algoritmineutraalisuuden, digiajan osallistavuuden (e-inclusion) ja yksityisen datan käytön haasteisiin. Kirjoittajien mukaan algoritmineutraaliuus tarkoittaa, että algoritmien päätöksentekoprosessien tulee olla läpinäkyviä ja ottaa ihmisten moninaisuus huomioon. Datavinoumien ehkäiseminen on myös tärkeä osa algoritmineutraalisuutta. Digiajan osallistavuudella tarkoitetaan, että tekoälyä tulee kehittää ja hyödyntää tukemaan yhteisöllisyyttä. Osallistavuus edellyttää kansalaisten laajamittaista osallistamista jo teknologian kehitysvaiheeseen ja digitaitojen kouluttamista. Villani -raportin kirjoittajat ajattelevat Tekoälyajan työ -raportin kirjoittajien tapaan, että työelämämuutoksien tutkimukseen panostaminen on yksi tärkeä osa digiajan osallistavuutta. Yksityisen datan käytön haasteet liittyvät puolestaan muun muassa yksilön suojan ja oikeuksien varmistamiseen tekoälyaikakautena, jolloin ihmisistä kerätään suuria määriä dataa tekoälyavusteista päätöksentekoa varten. Yksilöillä on lainmukainen oikeus saada tietää ja estää yksilöintiä varten kerättävää ja hyödynnettävää dataa, mutta ei yksilöistä kerättävää massadataa,¹⁶⁹ jota hyödynnetään esimerkiksi lainan myöntämisprosesseissa tai poliisipartioiden kohdentamisessa. Tekoälyavusteinen päätöksenteko ei ole läpinäkyvää yksilöille.¹⁷⁰

Ranskan ja Suomen hallitusten työryhmien ajatukset tekoälykehityksen vaatimuksista ovat hyödyntämisen etiikan mukaisesti rakennettu ja ne heijastavat holarkian mukaista ajattelua. Minkälainen kehitys on kokonaisuuden kannalta toivottavaa, on heille tärkeämpi

¹⁶⁷ Clark et al. 2018, 4; Koski & Husso 2018, 22; Villani et al. 2018, 11, 13, 19, 132.

¹⁶⁸ Koski & Husso 2018, 26, 49.

¹⁶⁹ Käytän tässä massadataa kuvaamaan suurta joukkoa dataa, jota hyödynnetään kokonaisuudessaan tilastollisten korrelaatioiden löytämiseen, mutta ei henkilöiden yksilöimiseen.

kysymys, kuin, minkälaista kehitystä voidaan saavuttaa.¹⁷¹ AI Sector Deal ja White House Summit on AI -raporttien kirjoittajat sanovat tosiaan myös, että tekoälykehityksen on tärkeä saavuttaa kansalaisten hyväksyntä. He kuitenkin perustelevat merkittävyyden sillä, että tekoälyteknologioiden käyttöaste on riippuvaista niiden yhteiskunnallisesta hyväksyttävyydestä.¹⁷² Näin ollen AI Sector Deal ja White House Summit on AI näkevät hyväksynnän välinearvona tekoälyn avulla saavutettavan tehokkuuden ja tuottavuuden saavuttamiseksi. Kansalaisten mielipiteen asettaminen välinearvoksi tehokkuuden ja tuottavuuden saavuttamiseksi kielii arvovinoumasta, jossa demokraattinen ajattelu palvelee talousjärjestelmää eikä toisin päin.

Ison-Britannian ja Yhdysvaltojen tekoälystrategia -työryhmien raportit ilmentävät Porterin ja Kramerin argumenttia, jonka mukaan länsimaiset taloustieteilijät ovat legitimoineet talousjärjestelmästä kapea-alaisen käsityksen, jonka mukaan maksimaalisen rahallisen voiton tavoittelu on paras keino saavuttaa kaikki inhimilliset ja yhteiskunnalliset tarpeet. Kapea-alainen näkemys tuottaa kuitenkin ulkoisvaikutuksia (*externalities*), joista kukaan ei ole laillisesti vastuussa. Kasvihuonepäästöt on hyvä esimerkki teollisten vallankumousten jälkeisen talousjärjestelmän ulkoisvaikutuksista.¹⁷³ Maija-Riitta Ollila, Erkki Laitila ja Yuval Harari argumentoivat, että vapaiden markkinoiden mekanismit tuottavat lyhyen tähtäimen päätöksiä, jotka johtavat valitsemaan taloudellisen tuottavuuden toiminnan perimmäiseksi tavoitteeksi. Tekoälyn hyödyntämisen haittavaikutukset, kuten algoritmivinoumat, voimistunut epätasa-arvo sekä yksilöiden oikeuksien heikentyminen voivat tulevaisuudessa olla kasvihuonepäästöjen tapaisia ulkoisvaikutuksia, jos kehityksen suunta jätetään vapaiden markkinoiden mekanismien varaan.

Tekoälystä koituvan hyödyn laaja-alainen jakautuminen ja kansalaisten kouluttaminen tekoälyn aikakautta varten yhdistetään kaikissa raporteissa kehityksen hyväksyttävyyteen. Hyödyn tasainen jakautuminen vaikuttaisi olevan raporttien mukainen tekoälyn etiikan minimivaatimus. Se sisältää ajatuksen hyvän tavoiteltavuudesta ja pahan (hyödyn jakautuminen vain harvoille) vältettävyydestä. AI Sector Deal ja White House Summit on AI -raporttien kirjoittajat sanovat, että tekoälyn tuottama hyöty, niin taloudellinen kuin palveluiden laadun paraneminen, jakautuu talouden mekanismien kautta. Näin ollen hallitusten tärkein tehtävä on varmistaa, että ne tarjoavat parhaat mahdolliset asetelmat

¹⁷⁰ Villani et al. 2018, 54, 86, 103, 114, 121, 132.

¹⁷¹ Koski & Husso 2018, 46; Villani et al. 2018, 13–14.

¹⁷² Clark et al. 2018, 3; OSTP 2018, 11.

¹⁷³ Andrew 2008; Porter & Kramer 2011, 4.

tekoälyteknologiaa kehittäville yrityksille, jotta nämä tuottaisivat taloudellista hyötyä ja palveluiden kehitystä juuri Isossa-Britanniassa ja Yhdysvalloissa.¹⁷⁴ Näkemys vahvistaa ennestään, että raporttien kirjoittajien näkemykset edustavat Porterin ja Kramerin esittämää kapean talousjärjestelmän käsitystä

Yksi tärkeä aihe houkuttelevuudessa on osaavan henkilöstön mahdollistaminen koulutuksen tai maahanmuuton avulla. White House Summit on AI -raportin työryhmä mainitsee väljästi, että työperäistä maahanmuuttoa voitaisiin helpottaa informaatioteknologian asiantuntijoiden osalta, ja että kansalaisten uudelleen kouluttamisen painopisteen tulisi olla kouluikäisissä. AI Sector Deal -raportin työryhmän mukaan koulutuksen olisi tärkeää kohdentua etenkin kouluikäisiin ja korkeakoulutettuihin ihmisiin. Näin varmistetaan, että Isossa-Britanniassa syntyy uusia osaajia ja potentiaaliset osaajat saavuttaisivat parhaan tasonsa. Lisäksi työryhmän mielestä on tarve miettiä, kuinka Iso-Britannia voisi houkuttaa tekoälyteknologian kansainvälistä osaamista.¹⁷⁵

Villanin ja Tekoälyajan työ -raporttien laatijoiden mielestä tekoälyn tuottamien hyötyjen laaja-alainen jakautuminen edellyttää hallitukselta aktiivista roolia. Talouden mekanismien varaan jätetyn tekoälykehityksen disruptiivinen vaikutus syventäisi yhteiskunnallista epätasa-arvoa, mikä voisi puolestaan johtaa yhteiskunnan koheesioon negatiivisesti. Työryhmien raporteissa sanotaan, että tekoälyn ymmärrystä syventävän koulutuksen tulisi pyrkiä kattamaan kaikki kansalaiset ja ottamaan huomioon ihmisryhmittymien erityistarpeita, joita voi olla esimerkiksi ikäryhmien väliset teknologiaosaamisen erot. Kattavalla koulutuksella pyrittäisiin ehkäisemään sitä, että tekoälyn disruptiivinen vaikutus muodostaisi väliinpuotoajaryhmiä. Tekoälyajan työ raportissa todetaan, että työpaikoilla tulee olemaan tärkeä rooli työikäisen väestön kouluttamisessa ja valtiolla sekä akateemisilla oppilaitoksilla on tärkeä rooli työpaikkojen tukemisessa. Raportin laatijat ehdottavat, että valtio tarjoaisi työikäisille suunnatun elinikäistä oppimista tukevan alustan, jonka muotoa ja tarjontaa kehitettäisiin yhteistyössä työnantajien ja työntekijöiden kanssa. Korkeakoulujen aloituspaikkojen tuntuva lisääminen mainitaan raportissa tärkeänä toimena koulutusluokkien poistamiseksi ja kattavan osaamisen varmistamiseksi.¹⁷⁶

Koulutuksen suuntaamisen lisäksi se, minkälaista koulutusta pidetään tärkeänä tekoälykehitykselle, erottaa valtioiden strategioita toisistaan. Ison-Britannian ja Yhdysvaltojen hallitusten työryhmät sanovat, että STEM-aineiden (science, technology,

¹⁷⁴ Clark et al. 2018, 3; OSTP 2018, 11.

¹⁷⁵ AI Sector Deal 2018, 16, 26; OSTP 2018, 5, 9.

engineering and maths) koulutus on keskiössä tekoälyaikakauteen valmistautumisessa, sillä ne ovat tärkeitä ohjelmoinnin ja sitä tukevien toimien osaamiselle. AI Sector Deal -raportin laatijoiden mukaan on myös tärkeä varautua eettisiin kysymyksiin, joita voi ilmetä tekoälyn kehittämisen ja hyödyntämisen myötä. Sen vuoksi työryhmä ohjeistaa Ison-Britannian hallitusta lisäämään STEM-aineiden lisäksi korkeakoulutusta erityisesti dataetiikassa. Raportin mukaisen dataetiikan olisi tarkoitus keskittyä datan turvallisen käytön varmistamiseen. Ison-Britannian hallituksen on määrä perustaa dataetiikan laitos valtiolliseen Alan Turingin yliopistoon.¹⁷⁷ Alan Turingin yliopisto on keskittynyt tietojenkäsittelytieteiden tutkimukseen, joten dataetiikan laitoksen ja muiden oppiaineiden välillä olisi mahdollisuus synergiaan. Suomen ja Ranskan hallitusten työryhmät painottavat, että poikkitieteellisen osaamisen varmistaminen on oikea tapa panostaa tekoälyn kehitykseen. STEM-aineet ovat vain yksi osa poikkitieteellistä kokonaisuutta, johon toivotaan lisäksi muun muassa etiikan, kognitiotieteiden, sosiaalitieteiden, taloustieteiden ja yhteiskuntatieteiden panosta. Villanin -raportissa suositellaan, että hallitus perustaisi poikkitieteellisiä tekoälytutkimusohjelmia eri puolilla maata sijaitseviin julkisiin korkeakouluihin.¹⁷⁸

Ison-Britannian ja Yhdysvaltain hallitusten työryhmät yhdistävät tekoälykehityksen kansantalouden kehitykseen suoremmin ja vahvemmin kuin Ranskan ja Suomen hallitusten työryhmät. AI Sector Deal ja White House Summit on AI -raporteissa mainitut koulutusjärjestelmämuutokset ilmentävät Martha Nussbaumin varoittamaa suuntausta, jossa kapea-alainen hyödyn tavoittelu korvaa laaja-alaisen sivistyksen. Rakennan argumenttini sille perustalle, että raporteissa ei puhuta STEM-aineiden painotuksen jaksottaisesta lisäämisestä, vaan STEM-aineiden merkitystä korostetaan, kuin muilla oppiaineilla ei olisi tekemistä tekoälykehityksen kanssa. Vaikka Ison-Britannian hallituksen työryhmä mainitsee dataetiikan tärkeänä aiheena, niin sekin lopulta tähtää alustatalouden päätarpeen, eli datan, vapauttamiseen tekoälyteknologioiden käyttöön. Työryhmä ymmärtää datan eettisen hyödyntämisen olevan tärkeä tekijä tekoälykehityksen legitimitetin kannalta.

AI Sector Deal, Villanin ja Tekoälyajan työ -raporteissa mainitaan, että STEM-aineiden koulutuksen houkuttavuuden lisäämisellä on mahdollisuus vaikuttaa tulevaisuuden tekoälyasiantuntijoiden monimuotoisuuteen. Raporttien kirjoittajat ajattelevat, että monimuotoisuuden tukeminen on tärkeää, koska ohjelmointi on tähän asti ollut selvästi miesvoittoinen ala. Villanin -raportissa sanotaan, että asiantuntijoiden monimuotoisuuden

¹⁷⁶ Koski & Husso 2018, 38–41, 49; Villani et al. 2018, 63, 94, 98.

¹⁷⁷ Clark et al. 2018, 7; OSTP 2018, 5.

myötä tekoälyn kehitys ottaisi paremmin huomioon eri ihmisryhmien tarpeet.¹⁷⁹ Näkemys antaa viitteitä siitä, että ohjelmoijien kokemusmaailmojen nähdään määrittävän, minkälainen ohjelmoinnin ajatellaan olevan merkityksellistä.

4.2 Eettisyys kestävässä pohjana

Erkki Laitila ja Maija-Riitta Ollila ehdottavat kestävä kehityksen asettamista tekoälyn hyödyntämisen perimmäiseksi tavoitteeksi. Laitila huomauttaa, että kestävä kehitys tavoittelee koko yhteiskunnallisen ja globaalin ekosysteemin hyvinvointia, jolloin sen toteuttamisessa ei ole varaa osaoptimointiin. Ekosysteemi tarkoittaa kestävä kehityksen kontekstissa yhtä lailla eliöiden ja elottomien ympäristötekijöiden muodostamaa ekologista yksikköä kuin toiminnallisten ihmisryhmittymien, eli organisaatioiden muodostamaa vuorovaikutuksellista yksikköä.¹⁸⁰ Laitilan ja Ollilan hyödyntämät määritelmät kestävästä kehityksestä rakentuvat niin sosiaalisesta kuin ekologisestakin hyvinvoinnista.

Ison-Britannian, Ranskan ja Suomen hallitusten työryhmien raporteissa todetaan, että tekoälykehityksen tulee pohjautua etenkin sosiaalisesti kestäväälle ajattelulle. Kyseisten työryhmien fokuksot eivät ole vain lähitulevaisuuteen reagoinnissa, vaan kehityksen pitkäaikaisten vaikutusten ennakkoinnissa. Ison-Britannian hallituksen työryhmä yhdistää sosiaalisen kestävyuden tekoälyteknologiassa hyödynnettävän datan eettisten kysymysten huomioimiseen. Työryhmän raportin mukaan etenkin herkkäluonteisen datan käytön riskit tulisi minimoida ennakkoon. Kansalaisten yksityisen datan vapautuminen tekoälyteknologioiden käyttöön mahdollistaisi julkisten ja yksityisten palveluiden laadun parantumisen ja halventumisen. Palveluiden laadun parantumista perustellaan sillä, että kansalaisten yksilölliset tarpeet kyettäisiin ottamaan paremmin huomioon palveluihin ohjaamisessa ja niiden muotoilemisessa.¹⁸¹

Yksilöllisten tarpeiden nykyistä parempi huomioon ottaminen esimerkiksi hoitoalalla on ehdottomasti tärkeä kehitysaskel inhimillisen hyvinvoinnin kohentamisessa. Siitä huolimatta se, että dataetiikan painopiste kaventuu datan saavutettavuuden dilemmaksi AI Sector Deal -raportissa, muistuttaa Ollilan problematisoimaa yritysten dataetiikkaa. Se on suhteellisen kapea näkemys dataetiikasta, tekoälyn etiikasta puhumattakaan.

Ranskan hallituksen työryhmän raportissakin mainitaan, että yksityisen ja yrityskohtaisen datan vapauttaminen tekoälyteknologioiden laaja-alaiseen käyttöön on tärkeä

¹⁷⁸ Koski & Husso 2018, 17, 38; Villani et al. 2018, 63.

¹⁷⁹ Clark et al. 2018, 16; Koski & Husso 2018, 39, 51; Villani et al. 2018, 133.

¹⁸⁰ Laitila 2019, 49; Ollila 2019, 22, 27.

askel kehityksen konkretisoimiseksi. Työryhmän mukaan datan vapauttamista varten olisi luotava kansalliset ja EU:n sisäiset säännöt datan käytölle ja säilyttämiselle. Suomen ja Ranskan hallitusten työryhmien näkemykset tekoälyn eettisistä vaatimuksista eivät kuitenkaan rajoitu dataetiikan pariin. Arvoja, joita Ranskan ja Suomen hallitusten työryhmien raporteissa korostetaan ovat tehokkuuden ja tuottavuuden kasvu, toiminnan avoimuus, luotettavuus, oikeudenmukaisuus, ja läpinäkyvyys. Tekoälyajan työ -raportissa mainitaan lisäksi, että Suomen hallituksen on hyvinvointivaltion edustajana tärkeä tunnistaa marginaaliryhmien tarpeet ja oikeudet tekoälyn hyödyntämistä suunniteltaessa.¹⁸²

Toiminnan luotettavuudella, avoimuudella ja läpinäkyvyydellä tarkoitetaan muun muassa sitä, että tekoälyn kehityksessä pyritään etsimään vastauksia *mustan laatikon*¹⁸³ ongelmalle, ja että myös kansalaisille on luotava mahdollisuus vaikuttaa tekoälyn kehitykseen ja hyödyntämiseen. Vaikuttamismahdollisuuksia pyritään parantamaan yhteiskunnan läpileikkaavan tekoälykoulutuksen avulla ja järjestämällä avoimia keskustelutilaisuuksia tekoälystä, joissa kehitystä koskevien ajatusten kuulluksi tuleminen ei ole riippuvainen puhujan yhteiskunnallisesta statuksesta. Oikeudenmukaisuus liitetään tekoälystä koituvan hyvän tasaiseen jakautumiseen, minkä ajatellaan vaativan hallituksen aktiivista roolia. Esimerkiksi Tekoälyajan työ -raportissa sanotaan, että ”laajan yhteiskunnallisen hyödyn tulisi olla kaiken julkisen toiminnan perusta”.¹⁸⁴ Villanin -raportin vastaava toteamus on, että tekoälykehitykseen vaikuttavien poliittisten päätösten tulee palvella yhteiskunnan koheesiota kahdella tavalla. Ensimmäiseksi täytyy varmistaa, että tekoälyn hyödyntäminen ei aiheuta uudenlaista, tai vahvasta olemassa olevaa epätasa-arvoa. Toiseksi tekoälyn hyödyntämisen tulee aktiivisesti pyrkiä yhtenäistämään ihmisryhmiä keskenään.¹⁸⁵

Villanin -raportti korostaa ekologisen kestävyys ja tekoälykehityksen yhteyttä sosiaalisen kestävyys rinnalla. Sen laatijat ehdottavat toimia tekoälyteknologian aiheuttamien ilmastopäästöjen vähentämiseksi, ja tekoälyn hyödyntämistä ilmastopäästöjen vähentämiseksi muilla aloilla. Tekoälyteknologian ilmastopäästöjen vähentämisen tarvetta perustellaan sillä, että informaatioteknologian päästöjen arvioidaan tuottavan 20–50% kaikista globaaleista päästöistä vuonna 2030. Raportin laatijoiden mukaan Ranskan hallituksella on velvollisuus pyrkiä pienentämään tekoälyteknologian ilmastovaikutuksia. Raportin

¹⁸¹ Clark et al. 2018, 9; Koski & Husso 2018, 45, 48; Villani et al. 2018, 6, 11, 14.

¹⁸² Koski & Husso 2018, 49; Villani et al. 2018, 132, 142.

¹⁸³ Mustan laatikon dilemmalla kuvataan tilannetta, jossa ihmisellä ei ole kykyä ymmärtää, miten tekoäly on päättänyt johtopäätöksensä.

¹⁸⁴ Koski & Husso 2018, 49.

¹⁸⁵ Villani et al. 2018, 132, 142.

kirjoittajien mukaan tavoitteeseen pyritään pääsemään kehittämällä tekoälyn ja ilmastomuutoksen suhdetta tutkiva laitos, sekä tukemalla eurooppalaisten pilvipalveluteknologioiden siirtymistä energiatehokkaammaksi. Näkemys heijastaa kokonaisvaltaista lähestymistapaa hyödyntämisen velvollisuuksiin. Jos valtio aikoo hyödyntää tekoälyä, tulee sen myös huolehtia, että hyödyntämisestä ei koidu ulkoisvaikutuksia.¹⁸⁶

Suomen ja Ison-Britannian hallitusten työryhmien raportit eivät ota yhtä laajasti kantaa ympäristökysymykseen kuin Villanin -raportti. AI Sector Deal -raportissa mainitaan, että tekoälystrategian lisäksi Ison-Britannian hallituksella on *puhtaan kasvun (clean growth)* strategia. Strategioiden erottaminen ei ota tekoälykehityksen ja ekologisen kestävyysvälistä synergian mahdollisuutta yhtä vahvasti huomioon, kuten Villanin -raportti tekee.¹⁸⁷ Tekoälyajan työ -raportin laatijat eivät puolestaan mainitse ilmastomuutosta lainkaan. Suomen hallituksen tekoälyohjelma ottaa kuitenkin ilmastomuutokseen kantaa toisessa raportissaan *Suomen tekoäly aika, Suomi tekoälyn soveltamisen kärkimaaksi: Tavoite ja toimenpidesuosituks* (2017), jossa mainitaan, että suomalaisilla yrityksillä on mahdollisuus vaikuttaa päästöintensiivisten teollisuuksien, kuten energia-alan, liikenneteollisuuden ja biotalouden muuntamiseksi ympäristöystävällisempään suuntaan tekoälyn hyödyntämisen avulla. Raportin laatijat perustelevat lähestymistavan positiivisia ilmastovaikutuksia sillä, että suomessa on korkean luokan osaamista ja hyvät datavarannot kyseisillä teollisuuden sektoreilla. Tekoälyn hyödyntämisellä olisi otollinen mahdollisuus tehostaa alojen prosesseja, luoda uusia innovaatioita ja pienentää alojen toiminnasta aiheutuvia ilmastopäästöjä.¹⁸⁸

4.3 Strategioiden ekosysteemit

Antroposeenin etiikka asettaa velvoitteen pohtia tekoälyn kehitystä ja hyödyntämistä kestävä kehityksen kautta. Kestävä kehitys puolestaan velvoittaa laajentamaan päämäärien tarkastelun yli kansallisten rajojen. Maija-Riitta Ollila kommentoi kirjassaan Suomen hallituksen tekoälyohjelmaa. Hän sanoo, että kansallisen kilpailukyvyn kasvattamiseen suuntaava ohjelma ohjaa päämäärät suuntautumaan sisäänpäin.¹⁸⁹ Tässä kohden Ollila on osittain väärässä. Tekoälyajan työ ja Villanin -raportit korostavat kansainvälisen hallitusten välisen yhteistyön tarvetta. Raporttien kirjoittajat sanovat, että se on tekoälykehityksen kannalta elintärkeää. He eivät tarkoita ainoastaan taloudellista yhteistyötä, vaan tekoälykehityksen kokonaisvaltaiseen yhtenäistämiseen pyrkivää yhteistyötä. Muun muassa valtioiden rajat

¹⁸⁶ Villani et al. 2018, 103.

¹⁸⁷ Clark et al. 2018, 6.

¹⁸⁸ Tekoälyohjelman ohjausryhmä 2017, 24.

ylittävä datavarantojen jakaminen ja datan käytön periaatteiden yhtenäistäminen Euroopan tasolla mainitaan raporttien toimintasuosituksissa. Suomen hallituksen työryhmän kokoonpanokin osoittaa, että kansainvälinen yhteistyö nähdään synergiaetuna. Suomen tekoälyohjelman työryhmää johtaa Pekka Ala-Pietilä, joka johtaa myös EU:n tekoälyohjelman työryhmää. Suomen ja Ranskan hallitusten työryhmien raporteista voi havaita hyvinkin samanlaisia suuntaviivoja, joita ovat muun muassa yhtenäisten arvojen kautta johtaminen ja hallituksen aktiivinen rooli kehitystä ohjaavana tahona. Suuntaviivat ovat nähtävissä myös EU:n ohjelmassa. Valtioiden ja kansainvälisen tason tekoälyohjelmia ei kannata leimata ennakkoon kapeakatseisiksi.¹⁹⁰

White House Summit on AI -raporttia voisi toisaalta käyttää vahvistamaan Ollilan argumenttia. Raportissa yhteistyöstä puhutaan sotatermistöin. Sen laatijat sanovat muun muassa, että Yhdysvaltojen kehitykselle on hyödyllistä, että sillä on liittolaisia (*allies*), joiden kanssa voidaan tehdä yhteistyötä tekoälyn tutkimus- ja kehitystoiminnassa. Liittolaisina mainitaan Ison-Britannian ja Ranskan hallitukset. Lisäksi raportissa korostetaan Yhdysvaltain hallitusta kansainvälisten neuvotteluiden johtohahmona. Siinä sanotaan, että Valkoisen Talon edustus neuvotteli yksinään innovaatioministereiden yhteiset linjat¹⁹¹ tekoälyn kehitykselle vuonna 2018 pidetyssä G7 maiden välisessä tekoälykonferenssissa. Neuvottelutilanne sisältää kuitenkin aina kahden tai useamman toimijan välistä vuorovaikutusta ja muiden toimijoiden mainitsematta jättäminen on oman toiminnan pönkittämistä. Raportissa luodaan lisäksi vastakkainasettelun retoriikalla mielikuva jaosta, jonka mukaan Yhdysvallat kuuluu sisäpiiriin ja muut valtiot sen ulkopuolelle. Retoriikan yksi tarkoitus voi olla yhdistää yhdysvaltalaisen koheesion tunnetta. Tärkeä huomio on, että liittolaisetkaan eivät kuulu Yhdysvaltojen sisäpiiriin, vaan heistä puhutaan erillisenä kokonaisuutena.¹⁹² White House Summit on AI -raportin laatijoiden päämäärät ovat tosiaan sisäänpäin kääntyneet ja kansainvälistä yhteistyön tarvettakin perustellaan vain Yhdysvaltain kansantalouden kautta.

Ison-Britannian hallituksen työryhmä ei käytä raportissaan vastakkainasettelun retoriikkaa, vaan työryhmän mukaan avoin kansainvälinen yhteistyö on tärkeää Ison-Britannian kehitykselle. Työryhmä sanoo AI Sector Deal -raportissaan, että Ison-Britannian

¹⁸⁹ Ollila 2019, 92.

¹⁹⁰ Ethics guidelines for trustworthy AI 2019; EU Member States sign up to cooperate on Artificial Intelligence 2018; Koski et al. 2018, 49; Tekoälyohjelman ohjausryhmä 2017; Villani et al. 2018, 20.

¹⁹¹ White House Summit on AI raportin kirjoittajat tarkoittavat tekoälykehityksen linjoilla kehityksen rahoitusta, tutkimus- ja kehitystoimintaan panostamista sekä tekoälyn soveltamisen parhaiden käytäntöjen jakamista hallitusten kesken. Parhailla käytännöillä tarkoitetaan tekoälyn hyödyntämistä eri teollisuussektorien tehokkuuden kasvattamiseksi.

¹⁹² OSTP 2018, 2–3, 11.

pitää pyrkiä houkuttelemaan kansainvälisiä tekoälyteknologiayrityksiä perustamaan toimipisteitään Isoon-Britanniaan. Työryhmä painottaa, että on yhtä lailla tärkeää tukea isobritannialaisten tekoälyteknologiayritysten kansainvälistymistä. Ollilan argumentin kannalta on tärkeä huomioida, että kyse ei ole hallitusten välisestä yhteistyöstä, vaan kansainvälisten yritysten ja Ison-Britannian hallituksen.¹⁹³ Yritystoiminnan merkityksen korostaminen painottaa, että työryhmän strategian päätavoite on varmistaa Ison-Britannian talouden kasvaminen. Toisaalta Brexit prosessi on ollut jo raportin laatimisen aikaan käynnissä. Voi olla, että hallituksen työryhmä on nähnyt tarpeelliseksi keskittyä kansallisen strategian rakentamiseen eikä hallitusten välisen yhteistyön mahdollisuuden miettimiseen. Raportti ei heijasta negatiivistakaan suhtautumista valtioiden väliseen yhteistyöhön, vaan jättää aiheen neutraalisti mainitsematta.

Kestävän kehityksen mukainen toiminta pohjautuu globaalin perspektiivin lisäksi pyrkimykseen saada aikaan yhteiskunnan läpileikkaavaa vaikutusta. CSV:n ekosysteemi tähtää siihen, että kehitystoiminta olisi yhteiskunnan sektorit läpileikkaavaa. Tämä tarkoittaa, että julkishallinnon, kansalaisten, yritysten, kansalaisjärjestöjen ja korkeakoulujen edustajia osallistetaan kehitystyön suunnitteluun ja toteutukseen. Yhdysvaltojen hallituksen työryhmän raportissa ekosysteemillä tarkoitetaan puhtaasti taloudellisen kehityksen ekosysteemiä. Raportissa mainitaan, että valtio ja yritykset voivat yhdessä rahoittaa tekoälyn tutkimus- ja kehitystoimintaa. Kansalaisista puhutaan kuluttajina, joita varten tekoälyä kehitetään, ja työvoimana, jolle tarjotaan koulutusta STEM-aineissa. Raportin kirjoittajat sanovat, että Yhdysvaltain hallitus tulee perustamaan tekoälyvaliokunnan, jonka tärkein tehtävä on koordinoida valtiollista tekoälyn tutkimus- ja kehitystoimintaa. Valiokunta koostuu teknologiseen kehitykseen perehtyneistä ministereistä.¹⁹⁴ White House summit on AI -raportti osaoptimoi tekoälyn hyödyntämisen palvelemaan talousjärjestelmää. Raportin laatijat perustelevat näkökantaansa sillä, että kansantalouden hyvinvointi tarkoittaa laaja-alaista hyvinvointia. Perustelu on hyvä esimerkki Hararin huomautukselle, että liberaaleissa valtioissa taloudellinen vauraus yhdistetään edelleen vahvasti ihmisten hyvinvointiin. Kansantalouden kasvu ei kuitenkaan automaattisesti tarkoita koko kansan hyvinvoinnin kasvua, etenkin kun kapitalistinen talousjärjestelmä aiheuttaa nykymuodoissaan varallisuuden epätasaista jakautumista.¹⁹⁵

¹⁹³ Clark et al. 2018, 20, 32.

¹⁹⁴ OSTP 2018, 3, 7, 9.

¹⁹⁵ Weeden & Brusky 2013.

Ison-Britannian, Ranskan ja Suomen hallitusten työryhmien raportit rakentavat kokonaisvaltaisemman kuvan yhteiskunnan ekosysteemistä kuin Yhdysvaltojen hallituksen työryhmä. Villanin ja Tekoälyajan työ -raportit kuvaavat, kuinka hallitustoiminnan tulisi pyrkiä varmistamaan, että tekoälykehitys ei pelkästään ota huomioon kansalaisia, vaan että kansalaiset ovat tärkeä osa kehitystä ohjaavaa kokonaisuutta. Tämän toteutuminen edellyttää, että kansalaiset ovat kykeneviä keskustelemaan tekoälyn kehityksen ja hyödyntämisen tavoitteista ja tavoista. Lisäksi työryhmät painottavat tarvetta järjestää kansalaisille mahdollisuus saada äänensä kuuluviin tekoälyä koskevissa yhteiskunnallisissa keskusteluissa.¹⁹⁶ AI Sector Deal -raportti ei suoranaisesti ota kansalaisia huomioon kehitystä ohjaavina tahoina, mutta raportin mukaan heidän tarpeensa ohjaavat kehitystä. Esimerkiksi yksityisyyden suojan varmistaminen dataetiikan toimintona on tärkeä sen vuoksi, että kansalaiset pitävät sitä tärkeänä.¹⁹⁷ Tässä mielessä Villanin ja Tekoälyn ajan työ -raporttien laatijoiden voidaan sanoa määrittävän kansalaiset Erkki Laitilan tapaan kehitykseen osallistuviksi subjekteiksi, ja AI Sector Deal -raportin laatijoiden voidaan puolestaan sanoa määrittävän kansalaiset kehitykseen myötävaikuttaviksi objekteiksi.

Ison-Britannian, Ranskan ja Suomen hallitusten työryhmien mukaan tekoälyn kehitystä monitoroivan ja ohjaavan julkisen elimen perustaminen kuuluu hallitusten strategioihin. Elinten on tarkoitus toimia linkkinä julkisen hallinnon ja muiden tekoälyn kehityksestä vastaavien tahojen välillä. AI Sector Deal -raportin laatijat sanovat, että Ison-Britannian hallitukseen perustetaan tekoälyministerin virka ja sitä ohjeistava tekoälyneuvosto. Neuvosto koostuu hallituksen, korkeakoulujen ja yritysjohtajien asiantuntijoista. Neuvosto ohjaa tekoälystrategian käyttöön ottoa niin teollisuudessa kuin valtion hallinnossakin.¹⁹⁸ Villanin -raportin mukaan Ranskan hallitus perustaa digitalisaatio- ja tekoälyteknologian eettisen valiokunnan, joka järjestää julkisia väittelyitä tekoälyn kehityksestä ja hyödyntämisestä. Valiokunnan on määrä myös ohjeistaa hallitusta tekoälyä koskevissa päätöksenteoissa.¹⁹⁹ Tekoälyajan työ raportin mukaan Suomen hallitus perustaa parlamentaarisen seurantaryhmän, jonka tavoite on ”edistää tekoälyn eettistä arvopohjaa”²⁰⁰ ja toimia asiantuntijan roolissa tekoälyn etiikan pilottihankkeissa ja tekoälyavusteisen päätöksenteon vastuukysymyksissä.

Raporteissa esitetyt näkemykset tekoälykehitystä monitoroivan ja ohjaavan elimen toiminnasta ja kokoonpanosta kertovat paljon raporttien laatijoiden ekosysteemikäsityksistä.

¹⁹⁶ Koski & Husso 2018, 25, 48, 50; Villani et al. 2018, 7, 142.

¹⁹⁷ Clark et al. 2018, 3, 9.

¹⁹⁸ Clark et al. 2018, 3, 10, 20.

¹⁹⁹ Villani et al. 2018, 16, 128.

AI Sector Deal -raportin kirjoittajat yhdistävät eri sektorien tekoälyn huippuosaaajia ratkomaan yhteiskunnallisia haasteita ja kansalaisille ei anneta lainkaan roolia. White House Summit on AI -raportin kirjoittajien mukaan tekoälyvaliokunnan tärkein tehtävä on koordinoida innovointia. Villanin ja Tekoälyajan työ -raporttien mukaan elinten on tarkoitus osallistaa kansalaisia tekoälykehitykseen ja tukea tekoälyn eettistä hyödyntämistä.²⁰¹ Yhdysvaltain hallituksen näkemys tekoälyvaliokunnasta on osoitus hallituksen tekoälykehityksen kapea-alaisesta ja osaoptimoivasta näkemyksestä. Ison-Britannian hallituksen ajatus tekoälyneuvostosta osoittaa, että tekoälykehityksen yhteiskunnalliset vaikutukset ymmärretään laaja-alaisesti, mutta sitä johdetaan teknokraattisesti. Ranskan ja Suomen hallitusten ajatukset eettisestä valiokunnasta ja parlamentaarisesta seurantaryhmästä edustavat raporteista laajinta näkemystä tekoälyn ekosysteemistä, sillä niiden on tarkoitus avustaa laaja-alaisen eettisten kysymysten ratkaisemisessa ja ne osallistavat kansalaiset osaksi toiminnallista ekosysteemiä. AI Sector Deal ja White House Summit on AI -raportit edustavat työntöohjauksen mukaista, ja Villanin sekä Tekoälyajan työ -raportit imuvetoista tapaa ohjata tekoälykehitystä.

4.4 Eettisyys kehityksen hidasteena

Kaikissa raporteissa mainitaan, että liiallisella sääntelyllä hallitus voi aiheuttaa aivovuotoa, mikä pienentäisi tekoälyn kehityksestä koituvia yhteiskunnallisia hyötyjä. Ison-Britannian, Suomen ja Ranskan hallitusten työryhmien raporttien laatijat eivät kuitenkaan ajattele, että ohjaavat toimenpiteet, kuten esimerkiksi tiedostettujen haitallisten toimien poissulkeminen sääntelyn avulla ja rahoituksen kohdentaminen tiettyjen eettisten ohjenuorien mukaisesti aiheuttaisivat samanlaista vaaraa.²⁰² Yhdysvaltojen hallituksen työryhmän raportti eroaa tässä suhteessa muista vertailtavista raporteista.

White House Summit on AI -raporttia tulee käsitellä suhteessa Yhdysvaltain vuoden 2016 hallituksen tekoälyohjelman kolmiosaiseen luonnokseen, sillä valmistellun ohjelman sivuun siirtäminen on osoitus peräkkäisten hallitusten välisestä erimielisyydestä. Vuonna 2016 valmistuneen luonnoksen tehnyt työryhmä kuvasi tekoälyn kehityksen haasteellisempänä kuin White House Summit on AI -raportin työryhmä. Heidän mukaansa hallituksen tulisi varautua puuttumaan lainsäädännön ja rahoituksen kohdentamisen keinoin kehityksen suuntaan. Visio oikeasta suunnasta tulisi pohjautua arvoille, joita tekoälyn avulla

²⁰⁰ Koski & Husso 2018, 50.

²⁰¹ Clark et al. 2018, 10–11; Koski & Husso 2018, 50; OSTP 2018, 7, 9; Villani et al. 2018, 63, 86, 128.

²⁰² Clark et al. 2018, 18; Koski & Husso 2018, 51; Villani et al. 2018, 76, 128.

halutaan vahvistaa ja tietoiselle riskien ehkäisemiselle. Luonnos korosti oikeudenmukaisen, reilun ja vastuullisen kehityksen merkitystä tekoälyn valtiollisessa hyödyntämisessä.²⁰³

White House Summit on AI -raportissa sanotaan, että yhteiskunnallisen hyväksynnän saavuttaminen on tärkeää, mutta hallitus ei lähde taistelemaan ”mielikuvitushirviöitä”²⁰⁴ (*imaginary beasts*) vastaan. Tämän jyrkän metaforan esittämisen jälkeen raportin laatijat kertovat, kuinka Wrightin veljesten ja Isaac Newtoninkaan keksintöjä ei pyritty etukäteen rajoittamaan lainsäädännön turvin. Wrightin Veljesten ja Newtonin keksintöjen ja tekoälyn yhteyttä ei selvitetty tarkemmin kuin vain sanomalla, että liika hallinnollinen ennakointi tyrehdyttää mahdollisuuden innovointiin. Raportissa ei sanota suoraan, että nämä kohdat olisivat vastaus edellisen hallituksen ohjelmaluonnokselle, mutta ilman yhteyttä luonnokseen mielikuvitushirviö -metaforan käyttö jää vaille merkitystä. Samoin sitä edeltävä toteamus, että yhteiskunnallinen hyväksyttävyys on toki tärkeää, osoittaa, että tekstillä viitataan johonkin aiemmin esitettyyn ajatukseen tekoälyn kehityksen hallinnoinnista.

Yhdysvaltain hallituksen työryhmä puhuu tekoälystä moraalisesti neutraalina asiana. Ollila, Harari ja Laitila pyrkivät juuri tämän kaltaisen ajattelun haastamiseen. Ajatus, että tekoälyn kehitystä ohjaa talouden näkymätön käsi, piilottaa tekoälyn kehityksen inhimillisen toiminnan ulkopuolella tapahtuvaksi toiminnaksi. Yhdysvaltain hallituksen työryhmä ei tunnista myöskään holarkian ja CSV:n mukaista ajatusta, jonka mukaan tekoälykehityksen päämäärien yhtenäistäminen olisi kaikkien hyöty. Työryhmän tapa mieltää hallituksen rooli tekoälykehityksen passiivisena mahdollistajana saattaa kuitenkin koitua pitkällä aikavälillä Yhdysvaltain kehityksen kannalta enemmän haitalliseksi kuin hyödylliseksi. Tekoälyä koskevat lainsäädännöt voivat nimittäin luoda jotain, mihin yritykset eivät yksinään yllä. EU:n yleinen tietosuojalaki GDPR ja maksupalveludirektiivi PSD2 ovat hyviä esimerkkejä siitä, kuinka lainsäädännöllä voidaan parantaa kansalaisten oikeuksia ja luottamusta tekoälyn aikaisessa maailmassa.²⁰⁵ CSV:n kannalta ajateltuna, GDPR ja PSD2 ovat valtion hallitusten, yritysten asiantuntijoiden ja korkeakoulututkijoiden välisen yhteistyön tulosta, jotka toimivat kansalaisten luottamusta lisäävinä standardeina. Kun kansalaisille on kirjattu teknologian kehityksen huomioon ottavia oikeuksia lakiin, on heidän helpompi luottaa, että yritykset ja muut tekoälyteknologiaa hyödyntävät tahot toimivat vastuullisten periaatteiden mukaisesti.

Luottamuksen vahvistamiseksi tarvitaan kuitenkin myös tieto siitä, että lakien noudattamista seurataan. Mainitsin aiemmin, että Google ja Facebook ovat saaneet osakseen

²⁰³ NSTC 2016, 17, 30.

²⁰⁴ OSTP 2018, 10.

huonoa mainetta tietosuojavuotojensa vuoksi. Kummatkin yritykset ovat saaneet useiden miljoonien eurojen sakkoja GDPR:n nojalla, sen astuttua voimaan toukokuussa 2018. Ranskan kansallista tietoturvaa valvova komissio CNIL antoi googlelle 50 miljoonan euron sakot, ja Italian vastaava elin sakotti Facebookilta 8.9 miljoonaa punttaa (noin 10 miljoonaa euroa). Kyseiset summat eivät kaada Facebookin ja Googlen kaltaisia teknologiateollisuuden jättiläisiä, mutta ovat todennäköisesti tarpeeksi isoja taloudellisia menetyksiä ohjatakseen yritysten johtavia työntekijöitä ja osakkaita muuttamaan yritystoiminnan käytäntöjä GDPR:n mukaisiksi. Rahallisen menetyksen lisäksi GDPR:n nojalla tehtävät sakotukset nostavat sakotetut yritykset epämieluisaan valoon mediassa ja ne voivat kärsiä luottamuksen kadosta.²⁰⁶

Raporttien heijastaman eettisen pohdinnan vertailussa on korostunut, että Ison-Britannian ja Yhdysvaltojen hallitusten työryhmät asettavat tekoälystrategioiden perimmäiseksi tavoitteeksi kansantalouden kohentumisen. Ranskan ja Suomen hallitusten työryhmien asettamat strategioiden tavoitteet ovat kokonaisvaltaisempia. Ne sisältävät kestäväen kehityksen ja oikeudenmukaisuuden vaatimukset, joiden sivutuotteena syntyy myös taloudellista vaurautta. Vaikka argumentoin, että Ison-Britannian ja Yhdysvaltojen hallitusten työryhmien päämäärät ovat osiltaan yhteneviä, niin haluan korostaa, että AI Sector Deal -raportin päämäärä sisältää laajemman käsityksen hyödyntämisen etiikasta kuin White House Summit on AI -raportin päämäärä. Se käy ilmi, kun raporteissa esiteltyjä strategioita tarkastellaan etiikan aikaraamin näkökulmasta.

White House Summit on AI -raportin heijastama eettinen pohdinta johdattaa ajatukseen, että tekoälykehitys on itsessään moraalisesti neutraalia, ja että sen luomat dilemmat tulevat näkyväksi vasta kasvavan käytön myötä. Tekoälykehityksen potentiaali tuottaa taloudellista hyötyä on suuri ja kansantalouden kasvu tarkoittaa myös yhteiskunnallisen hyvinvoinnin kasvua. Hallituksen tulee omalla toiminnallaan mahdollistaa tekoälyn kehittäminen ja hyödyntäminen muun muassa poistamalla innovoinnille haitallisia lainsäädäntöjä, ja reagoida, jos haasteita ilmenee. AI Sector Deal -raportin heijastama eettinen pohdinta osoittaa, että yhteiskunnassa on tunnistettavissa arvoja, joita tekoälykehityksen tulee ottaa huomioon, jotta kehitys saavuttaa yhteiskunnallisen hyväksynnän. Arvoista merkittävimmät ovat yksilön suoja ja oikeudenmukaisen datan vaaliminen. Näiden arvojen toteutuminen tekoälyn aikakautena edellyttää tekoälykehityksen ennakoimista ja dataetiikan huomioimista teknologian

²⁰⁵ Yleinen tietosuoja-asetus 2019; Payment services (PSD2) s.a..

²⁰⁶ Hern 2018; Räisänen 2018; Hern 2019.

kehitysvaiheessa ja lainsäädännössä. Dataetiikka edellyttää tekoälyteknologiasta saatavien tämän hetkisten tietojen hyödyntämistä tavoiteltavien ja vältettävien tulevaisuusskenaarioiden luomiseksi.

Villanin ja Tekoälyajan työ -raporttien heijastama eettinen pohdinta asettaa tekoälykehityksen tavoitteeksi yhteiskunnallisen hyvinvoinnin. Raporttien laatijat eivät johda hyvinvointia tiettyjen yhteiskunnallisten järjestelmien kautta rakentuvaksi, vaan he hahmottavat kokonaisuutta yhdistäviä päämääriä, jotka ohjaavat sektoreita tekemään yhteistyötä. Kestävästä kehityksestä muokataan toimintaa ohjaavia periaatteita, kuten algoritmineutraalius ja marginaaliryhmien tukeminen, ja niiden toteuttamisen tarve argumentoidaan kokonaisuuteen vaikuttavina toimina. Tekoälyajan työ -raportissa sanotaan useaan otteeseen, että kehityksen suuntaa mietittäessä ei ole kyse siitä, mikä on teknologisesti mahdollista toteuttaa, vaan siitä, minkälainen kehitys on toivottavaa. Hallituksen tulee tukea yhteisten päämäärien muodostumista ja hahmottamista, sekä ohjata tekoälyn kehityksen ja hyödyntämisen toimintaa niitä kohti.

Hyödyntämisen etiikan näkökulmasta tarkasteltuna voidaan todeta, että White House Summit on Ai -raportti siirtää hyödyntämisen vastuun täysin tekoälyä kehittäville yrityksille. Raportin kirjoittajat eivät tunnista, että hallitus on yhtä lailla vastuussa tekoälykehityksestä kuin yritykset, kun se pyrkii kiihdyttämään tekoälyn kehittämistä ja hyödyntämistä ja saavuttamaan siten taloudellista hyötyä. Raportin kirjoittajat ovat valinneet ottaa vastaan hyödyntämisen tuotot, mutta eivät vastuuta.

AI Sector Deal -raportin kirjoittajat puolestaan tunnistavat, että tekoälykehitykseen osallistuminen tuottaa hallitukselle myös vastuita. Kirjoittajien tunnistamat vastuut ulottuvat kuitenkin lähinnä siihen, että hallituksen on varmistettava tekoälyavusteiseen päätöksentekoon hyödynnettävän datan turvallinen ja oikeudenmukainen käyttö. Dataetiikan ja osaksi ohjelmoinnin etiikan näkökulmasta raportin ehdottamat toimet ovat kattavia. Raportti ei kuitenkaan ilmennä laajoja hyödyntämisen eettisiä periaatteita, vaan siirtää hyödyntämiseen liittyviä vastuita White House Summit on AI -raportin tapaan vapaiden markkinoiden mekanismien varaan. Tämä ilmenee esimerkiksi siinä, että tekoälyn mahdollisiin disruptiivisiin vaikutuksiin ei oteta raportissa kantaa ja että tekoälyn hyötyjen ajatellaan leviävän markkinoiden mekanismien kautta palvelemaan koko yhteiskuntaa.

Tekoälyajan työ ja Villanin -raportit pyrkivät varmistamaan, että tekoälykehitys ei olisi teknokraattista vaan, että kehityksessä otettaisiin myös kansalaisten ajatukset huomioon. Se kuvastaa, että raporttien kirjoittajat tunnistavat, että hallituksilla on vastuu huolehtia, että tekoälykehitys hyödyttää koko yhteiskuntaa. Raporttien kirjoittajat tunnistavat myös, että

tekoälykehitys on globaali ilmiö, mikä asettaa vaatimuksen pyrkiä tiiviiseen kansainväliseen yhteistyöhön, jos tekoälykehitykseen halutaan vaikuttaa laajassa mittakaavassa. Edelleen raportit ilmentävät, että niissä toiminnan päämäärien valikoitumista ohjaavat pitkälti kysymykset, mitä arvoja toiminnalla halutaan saavuttaa ja miten valitut päämäärät vaikuttavat yhteiskuntaan kokonaisuudessaan. Kummankin raportin tavoitteiden taustaehtona on oikeudenmukaisuuden periaate, mikä välittyy erityisesti siinä, miten kummassakin raportissa on mietitty asioita paljon marginaaliin jäävien ihmisryhmien kautta.

5 Loppukatsaus

Yuval Hararin rakentamat visiot tulevaisuudesta, joissa ihmisen rooli kaventuu teknologian kehitystä tukevaksi, ilmentävät Ray Kurtzweilin ajatusten mukaista transhumanistista maailmankatsomusta. Sen ei voida sanoa edustavan suuren ihmisjoukon ajattelua. Hararin pelko kuitenkin on, että jos teknologiakehityksen tavoitteista ei käydä avointa ja laajaa keskustelua, sen kehitys tapahtuu piilossa suurimmalta osalta ihmisistä. Kurtzweil johtaa Googlen koneoppimista kehittävää tutkimusryhmää, mikä tarkoittaa, että hänen ajatuksillaan on painoarvoa yhden maailman suurimman tekoälyä kehittävän yrityksen toiminnassa. Se saa kysymään, millaisia visioita muilla vaikutusvaltaisilla tekoälyä kehittäville tahoilla on.

Analyysini perusteella valtioiden hallituksilla on hyvät mahdollisuudet vaikuttaa tekoälykehityksestä käytävään arvokeskusteluun. Hallituksilla on myös mahdollisuus vahvistaa yhteiskunnallisten sektorien yhteistyötä, jolloin tavoiteltavaksi katsottujen arvojen tavoittelu ja havaittujen riskien minimointi voidaan muuttaa koordinoituksi toiminnaksi. Koordinoinnin onnistuminen edellyttää holarkista yhteiskuntakäsitystä, mikä tarkoittaa, että tekoälyn kehityksellä pyritään koko yhteiskunnan kehitykseen ja että kehityksessä asetetaan erityistä painoarvoa yhteiskunnallisten sektorien väliselle vuorovaikutukselle.

Se, että elämme antroposeenissa asettaa ihmisille entistä selvemmin velvollisuuden tavoitella kestävästä kehitystä. Kestävä kehitys voidaan jakaa niin ekologiseen kuin sosiaaliseen kestävyyteen. Erkki Laitilan mukaan kestävä kehitys on erinomainen tavoite holarkian näkökulmasta, sillä sen toteutuminen edellyttää yhteiskuntasektorien yhteistyötä. Tähän asti sosiaaliset kehityshankkeet sekä tekoälyn kehitys ovat olleet hajautuneita. Creating Shared Value -teoria on hyvä työkalu edistää julkisen hallinnon, yritysten, korkeakoulujen, kansalaisjärjestöjen ja kansalaisten välistä yhteistyötä. Se edistää yhteisvoiton ajatusta ja tavoittelua ja mahdollistaa siten monimutkaisten sosiaalisten ongelmien ratkaisemista. Näin ollen se on tärkeä osa tekoälyn hyödyntämisen teoreettista viitekehystä.

Strategiaraporttien analyysin ja Maija-Riitta Ollilan, Erkki Laitilan ja Yuval Hararin näkemysten perusteella voidaan sanoa, että tekoälykehityksellä on potentiaalia muodostaa markkinahäiriöitä, jos sen kehitys jätetään vapaiden markkinoiden varaan. Valtiollisesti ja kansainvälisesti koordinoimaton vapaan markkinatalouden mekanismien ohjaama tekoälykehitys voi lähitulevaisuudessa aiheuttaa algoritmivinousia, epätasa-arvoisuuden voimistumista, yksilöiden oikeuksien heikentymistä, inhimillisen merkityksellisyyden katoa, varallisuuden epätasaisen jaon voimistumista ja ilmastonmuutoksen kiihtymistä.

Ison-Britannian, Ranskan, Suomen ja Yhdysvaltojen hallitusten tekoälystrategioita laatineiden työryhmien raporttien vertailussa ilmeni, että tekoälyn valtiollisen hyödyntämisen eettinen minimivaatimus voisi olla, että kansalaiset hyväksyvät hyödyntämisen. Tekoälystä koituvan hyödyn tasainen jakautuminen on raporttien mukainen kansalaisten hyväksynnän taustaehto. Käytännössä hyväksyntä on ainakin osittainen minimivaatimus tekoälyn hyödyntämiselle, sillä se on kehityksen legitimitietin ehto.

On kysyttävä, onko kansalaisten hyväksyntä riittävä kriteeri tekoälyn etiikalle. Vastaus riippuu siitä, miten kansalaisten hyväksynnän ajatellaan muodostuvan. Jos se nähdään Creating Shared Value -teorian mukaisesti yhteiseen hyvään pyrkivän toiminnan tuloksena, kuten Ranskan ja Suomen hallitusten työryhmien laatimat raportit kuvaavat, niin se sisältää suhteellisen kattavan eettisen periaatelistan. Sen keskeisiä periaatteista ovat oikeudenmukaisuus ja yhteiskunnallisen diskurssin mahdollistaminen. Kansalaisten kyky ja mahdollisuus osallistua tekoälyä koskevaan keskusteluun tulisi varmistaa koulutuksen avulla ja perustamalla kaikille avoimia tekoälyä käsitteleviä foorumeita. Kansalaisten puheille tulisi myös antaa painoarvoa kehityksen suuntaa valittaessa. Näin varmistetaan, että kansalaiset osallistetaan mukaan kehitykseen, jolloin he ovat tekoälykehitystä määrittäviä subjekteja eikä vain sen objekteja.

AI Sector Deal -raportin mukaan tekoälyä kehittävien ja hyödyntävien tahojen on varmistettava, että tekoälyavusteinen päätöksenteko ei sisällä vinoumia ja että herkkäluonteista dataa käsitellään turvallisesti. Näin tekoälykehitys ottaisi huomioon kansalaisten tarpeita ja oikeuksia. Raportin mukaan Ison-Britannian hallituksella on keskeinen rooli tukea edellä mainittujen haasteiden selvittämistä. Kummatkin toimet tähtäävät kuitenkin tekoälykehityksen oikeuttamiseen eivätkä sisällä laajaa käsitystä tekoälyn eettisestä hyödyntämisestä. Tekoälykehityksen oikeutus on myös tekoälykehityksen edellytys. Kun otetaan lisäksi huomioon, että raportin kirjoittajat toivovat tekoälyn tehostavan Ison-Britannian taloudellista kilpailukykyä ja Ison-Britannian hallituksen asemaa kansainvälisenä poliittisena toimijana, voidaan todeta, että AI Sector Deal -raportin ilmentämä perimmäinen motiivi tekoälyn etiikalle on talouskasvun varmistaminen ja suuremman poliittisen vaikutusvallan saavuttaminen.

Yhdysvaltain hallituksen työryhmän kuvaamat toimet, jotta tekoälyn kehitys ja hyödyntäminen saavuttavat yhteiskunnallisen hyväksyttävyyden, ovat epämääräisemmät kuin Ison-Britannian, Ranskan ja Suomen hallitusten. White House Summit on AI -raportin mukaan vapaan markkinatalouden mekanismit pitävät huolen siitä, että tekoälykehitys hyödyttää koko yhteiskuntaa. Todellisuudessa vapaiden markkinoiden mekanismit ovat

tuottaneet ja voimistaneet globaaleja markkinahäiriöitä, kuten ilmastonmuutosta ja suurien tuloerojen muodostumista. Vapaiden markkinoiden mekanismit luovat paineen lyhytaikaisten taloudelliseen tuottavuuteen tähtäävien tavoitteiden asettamiseksi, mikä on ristiriidassa tekoälyn eettisen hyödyntämisen edellytysten kanssa. Tekoälyn eettisen hyödyntämisen tulisi tähdätä ylisukupolviseen hyvinvointiin. Antroposeenin etiikka on hyvä teoreettinen viitekehys tekoälyn hyödyntämisen arvoille ja kestävä kehitys arvoista johdettaville periaatteille ja tavoitteille.

Hyödyntämisen etiikan toteutumisen edellytys on, että valtion hallinto tunnistaa erityisasemansa tekoälyn vuorovaikutteista kehitystä ohjaavana tahona. Valtion hallinto kykenee arvokeskusteluiden fasilitoinnin ja tukemisen, lainsäädännön, koulutuksen tarjonnan, kansainvälisten yhteistyösopimusten, rahoituksen kohdentamisen ja julkisten hankintojen sekä palveluiden tarjonnan kautta vaikuttamaan laaja-alaisesti CSV:n mukaisen ekosysteemin toimintaan. Erityisasemiensa vuoksi on tärkeää, että valtion hallinnot eivät jää passiivisiksi toimijoiksi tekoälyn etiikan kehittämisen osalta.

Stuart Armstrongin ja Nick Bostromin ajatus supertekoälyn kehityksen kilpajuoksusta tuntuu pätevän myös kansallisen tason tekoälystrategioihin. He esittävät huolensa siitä, että kun ohjelmoijien tavoite on olla ensimmäinen, joka kehittää supertekoälyn/vahvan yleisälyn, he eivät ota kehityksen riskejä tarpeeksi hyvin huomioon. Näin kilpajuokutilanne voi johtaa ihmisten ja ympäristön kannalta tuhoisiin lopputuloksiin. Hallitukset tavoittelevat supertekoälyn kehittämisen sijaan taloudellisen kilpailukyvyn kasvattamista tekoälyn avulla. Yhdysvaltain hallituksen työryhmän esittämä strategia on hyvä esimerkki kilpajuoksun sokeuttamisesta. Työryhmä ei kykene näkemään edes eettisen sääntelyn pitkän aikavälin taloudellisia hyötyjä, joiden kasvattaminen on työryhmän rakentaman strategian päätavoite. Ison-Britannian strategia osoittaa, kuinka kilpailuasetelma voi johtaa välinearvojen vääristymistä itseisarvon asemaan. Suomen ja Ranskan strategiat antavat puolestaan viitteitä siitä, että valtiollisia ohjelmia on mahdollista kehittää itseisarvojen kautta, vaikka yksi kehityksen tavoitteista olisi saavuttaa kasvavaa kilpailukykyä.

Keskityin tässä tutkielmassa tarkastelemaan valtioita tekoälyn hyödyntäjinä eettisestä näkökulmasta tarkasteltuna. Sen tulokset antavat lisää näkökulmia tekoälyn poliittisesta hallinnoinnista käytävään keskusteluun. Tutkielmani näkökulmaa voisi täydentää jatkotutkimuksella, jossa tarkastellaan useamman maanosan strategioita. Lisäksi strategioiden ilmentämiä eettisiä näkemyksiä voisi analysoida tarkasteltavien valtioiden poliittisen historian kautta. Näkökulman tulisi keskittyä teknologiakehityksen poliittiseen hallintaan. Historiallinen analyysi antaisi paremmat lähtökohdat ymmärtää, miten Ison-Britannian,

Ranskan, Suomen ja Yhdysvaltojen hallitusten työryhmien strategiat ovat muotoutuneet sellaisiksi kuin ne nyt ovat.

Olisi myös tärkeää tehdä jatkoselvitystä siitä, millaisiksi strategioiden käytännön toteutukset muodostuvat. Strategiat kun ovat kuitenkin suunnitelmia, jotka vasta pohjustavat toimintaa. Jatkotutkimus syventäisi ymmärrystämme tekoälyn eettisen hyödyntämisen keskeisistä haasteista ja mahdollisuuksista.

Yhdysvaltojen hallitus julkisti kattavamman raportin tekoälystrategiastaan sen jälkeen, kun olin saanut strategioiden vertailuni valmiiksi. Tulin tietoisesti kyseisestä raportista vasta palauttaessani tutkielmaani enkä sen vuoksi voinut huomioida sitä tämän enempää. Kyseisen raportin laatinut työryhmä on saattanut huomioida tekoälyn etiikan kysymyksiä tässä tutkielmassa hyödynnettyä White House Summit on AI -raporttia laajemmin.

Lähde- ja kirjallisuusluettelo

Käytetyt lyhenteet

PERRI: Public Engagement in Responsible Research and Innovation (s.a.).

OSTP: The White House Office of Science and Technology Policy (2018).

NSTC: Executive Office of the President, National Science and Technology Council, Committee on Technology (2016).

Lähteet ja apuneuvot

Clark, Greg, Matt Hancock, Dame Wendy & Jerome Pesenti (2018). Industrial Strategy – Artificial Sector Deal. *Publishing.go.uk*.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/702810/180425_BEIS_AI_Sector_Deal__4_.pdf (haettu: 8.12.2019).

Executive Office of the President, National Science and Technology Council, Committee on Technology (2016). *Preparing for the Future of Artificial Intelligence*.
https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (haettu: 8.12.2019).

French Strategy for Artificial Intelligence (2018). *AI For Humanity*.
<https://www.aiforhumanity.fr/en/> (haettu: 8.12.2019).

Glare, P. G. W. 1992, *Oxford Latin dictionary*. Repr. Oxford: Oxford University Press.

Koski, Olli & Kai Husso (toim.) (2018). Tekoällyajan työ – Neljä näkökulmaa talouteen, työllisyyteen, osaamiseen ja etiikkaan. *Työ- ja elinkeinoministeriön julkaisuja* 19/2018.

MOT Pro Englanti (s.a). *MOT*. Kielikone Oy & Gummerus Kustannus Oy.

Tekoälyohjelman ohjausryhmä (2017). Suomen tekoälyaika – Suomi tekoälyn soveltamisen kärkimaaksi: Tavoite ja toimenpidesuosituksset. *Työ- ja elinkeinoministeriön julkaisuja TEM raportteja* 41/2017.

The White House Office of Science and Technology Policy (2018). Summary of the 2018 White House Summit on Artificial Intelligence for American Industry. *Whitehouse.gov*. <https://www.whitehouse.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf> (Haettu: 8.12.2019).

Villani, Cedric, Marc Shoenauer, Yann Bonnet, Charly Berthet, Anne-Charlotte Corlut, Francois Levin & Bertrand Rondepierre (2018). For a Meaningful Artificial Intelligence – Towards a French and European Strategy. *AI for Humanity*.
https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf (haettu: 8.12.2019).

Kirjallisuus

Ackoff, Russel L. (1999). *Ackoff's Best*. New York: John Wiley & Sons.

- Abney, Keith, George A. Bekey & Patrick Lin (2012). *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, Mass.: MIT Press.
- Ahlroth, Jussi (2018). Addiktion algoritmi–Piilaakson pioneeri Jaron Lanier kertoo, miten Facebook ja Google ovat ottaneet mieleemme ja yhteiskuntamme valtaansa käyttäytymisen muokkauksella. *Helsingin Sanomat*. <https://www.hs.fi/nyt/art-2000005864675.html> (haettu 8.12.2019).
- Ailisto, Heikki (2018). Tekoälyn käsitekartta – VN -TEAS projekti Tekoälyn kokonaisjäsenitys ja kansallinen osaamiskartoitus. *VTT.fi*. https://www.vtt.fi/files/DataK%c3%a4sitekartta%20AI_2018.pdf (haettu: 8.12.2019).
- Airaksinen, Timo (1987). *Moraalifilosofia*, toinen painos. Helsinki: WSOY.
- Andrew, Brian. (2008). Market failure, government failure and externalities in climate change mitigation: The case for a carbon tax. *Public Administration and Development* 28:5 393–401. Doi:10.1002/pad.517.
- Aristoteles (2005). *Nikomakhoksen etiikka*. Suomentanut Simo Knuuttila. Helsinki: Gaudeamus.
- Armstrong, Stuart, Nick Bostrom & Carl Shulman (2013). Racing to the precipice: a model of artificial intelligence development. Technical Report#2013-1, *Future of Humanity Institute*, Oxford University: pp. 1–8. <http://www.fhi.ox.ac.uk/wp-content/uploads/Racing-to-the-precipice-a-model-of-artificial-intelligence-development.pdf> (haettu: 8.12.2019).
- Asikainen, Jaro (2018). Linnuiksi naamioidut dronekamerat vakoilevat taivaalta ihmisten liikkeitä – Kiinan muslimivähemmistön kontrolli on kuin Orwellin painajaisesta. *Yle*. <https://yle.fi/uutiset/3-10367198> (haettu: 8.12.2019).
- Asilomar Principles (2017). *Future of Life Institute*. <https://futureoflife.org/ai-principles/?cn-reloaded=1> (haettu: 8.12.2019).
- Asimov, Isaac (1950). Runaround. Teoksessa *I, Robot*. New York: Gnome Press. s.40.
- Augustinus, Aurelius (2003). Jumalan valtio: Osa 1 (kirjat 1–10). Suomentanut Heikki Koskeniemi. Helsinki: WSOY.
- Ayres, Robert U., Jeroen C. J. M. van den Bergh, & John M. Gowdy (1998). Viewpoint: Weak versus Strong Sustainability. *TI Discussion Paper*, no. 98–103/3. Amsterdam: Tinbergen Instituut (TI).
- Bennett, Nathan & G. James Lemoine (2014). What VUCA really means for you. *Harvard Business review* 92 (1/2). <https://hbr.org/2014/01/what-vuca-really-means-for-you> (haettu: 8.12.2019).

- Beauchamp, Tom L., ja James F. Childress (2013). *Principles of Biomedical Ethics*. 7th ed. New York: Oxford University Press.
- Bostrom, Nick (2009). The Future of Humanity. *Geopolitics, History, and International Relations*, Vol. 1, No. 2 (2009): 41–78.
- Bostrom, Nick & Eliezer Yudkowsky (2014). The ethics of artificial intelligence. Keith Frankish & William Ramsey (toim.), *The Cambridge Handbook of Artificial Intelligence* (pp. 316-334). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139046855.020.
- Bostrom, Nick & Anders Sandberg (2008). *The Wisdom of Nature: An Evolutionary Heuristic for Human Enhancement*. Human Enhancement Oxford University Press. 2008 Julian Savulescu & Nick Bostrom (toim.). <https://www.nickbostrom.com/evolution.pdf> (haettu: 8.12.2019).
- Bughin, Jacques, Jacob Staun, Jens R. Andersen, Martin Schultz-Nielsen, Peter Aagaard & Thyge Enggaard (2017). Digitally-enabled automation and artificial intelligence: Shaping the future of work in Europe's digital front-runners. *McKinsey & Company*. <https://www.mckinsey.com/~media/mckinsey/featured%20insights/europe/shaping%20the%20future%20of%20work%20in%20europes%20nine%20digital%20front%20runner%20countries/shaping-the-future-of-work-in-europes-digital-front-runners.ashx> (haettu: 8.12.2019).
- Brooker, Katrina (2015). Google Ventures and the Search for Immortality. *Bloomberg*. <https://www.bloomberg.com/news/articles/2015-03-09/google-ventures-bill-maris-investing-in-idea-of-living-to-500> (haettu: 8.12.2019).
- Chui, Michael, Martin Harryson, James Manyika, Roger Roberts, Rita Chung, Ashley van Heteren ja Pieter Nel (2018). Applying artificial intelligence for social good. *McKinsey Global Institute*. <https://www.mckinsey.com/~media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Applying%20artificial%20intelligence%20for%20social%20good/MGI-Applying-AI-for-social-good-Discussion-paper-Dec-2018.ashx> (haettu 8.12.2019).
- Cicmil, Svetlana & Damien Hodgson (2006). New possibilities for project management theory: A critical engagement. *Project Management Journal* 37(3), 111–122.
- Dutton, Jim (2018). An Overview of National AI Strategies. *Medium.com – Politics+AI*. <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd> (haettu: 8.12.2019).
- Elements of AI (s.a.). Helsingin Yliopiston ja Reaktorin verkkokurssi. *Elementsofai.com*. <https://www.elementsofai.com/fi/> (haettu: 8.12.2019).
- Ethics guidelines for trustworthy AI (2019). *Europa.eu – Digital Single Market*. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> (haettu: 8.12.2019).

- EU Member States sign up to cooperate on Artificial Intelligence (2018). *Europa.eu – Digital Single Market*. <https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence> (haettu: 8.12.2019).
- Facebook (s.a.). *Partnership on AI*. <https://www.partnershiponai.org/partners/facebook/> (haettu: 8.12.2019).
- Fact Sheet: Artificial Intelligence for Europe (2019). *Europa.eu – Digital Single Market*. <https://ec.europa.eu/digital-single-market/en/news/factsheet-artificial-intelligence-europe> (haettu 8.12.2019).
- Gibson, James J. (1979). Chapter 8 the Theory of Affordances. *The Ecological Approach to Visual Perception* Gibson, J. J.. Boston: Houghton Mifflin.
- Hallamaa, Jaana (2017). *Yhteistoiminnan Etiikka*. Helsinki: Gaudeamus.
- Hallamaa, Jaana (s.a.). Tutkielman laatimiseen liittyviä metodisia ja muodollisia näkökohtia. *Helsinki.fi/teologia*. http://www.helsinki.fi/teol/steol_vanha/opiskelu/Prosessi/Pros_raut10.html (haettu: 8.12.2019).
- Hansson, Sven O. 2017. *The ethics of technology: Methods and approaches*. London; New York: Rowman & Littlefield International, Ltd.
- Harari, Yuval N. (2017). *Homo Deus – Huomisen lyhyt historia*. Suomentanut Jaana Iso-Markku. Norhaven: Bazar.
- Heikkinen, Seppo (2018) Tekoäly muuttaa maailman – pian se tekee jopa lääkärin ja juristin töitä. *Yle.fi*. <https://yle.fi/aihe/artikkeli/2017/06/04/tekoaly-muuttaa-maailman-pian-se-tekee-jopa-laakarin-ja-juristin-toita> (haettu: 8.12.2019).
- Hern, Alex (2019). Google fined record £44m by French data protection watchdog. *The Guardian*. <https://www.theguardian.com/technology/2019/jan/21/google-fined-record-44m-by-french-data-protection-watchdog> (haettu: 8.12.2019).
- Hern, Alex (2018). Italian regulator fines Facebook £8.9m for misleading users. *The Guardian*. <https://www.theguardian.com/technology/2018/dec/07/italian-regulator-fines-facebook-89m-for-misleading-users> (haettu: 8.12.2019).
- Hern, Alex (2018). Google's solution to accidental algorithmic racism: ban gorillas. *The Guardian*. <https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people> (haettu: 8.12.2019).
- Hodgson, Damian & Svetlana Cicmil (2016). Making projects critical 15 years on: a retrospective reflection (2001-2016). *International Journal of Managing Projects in Business* 9. 744–751. DOI 10.1108/IJMPB-10-2015-0105.
- Hume, David (1998). *An enquiry concerning the principles of morals*. Suomentanut Tom L. Beauchamp. Oxford: Oxford University Press.

- IPCC (2018). *Summary for Policymakers. In: Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.* Toimittaneet V. Masson-Delmotte, P. Zhai, H. O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor & T. Waterfield. World Meteorological Organization, Geneva, Switzerland.
- Jolkkonen, Jari (2007). Systemaattinen Analyysi Tutkimusmetodina. *Joensuu.fi/teoltdk*.
https://courses.helsinki.fi/sites/default/files/course-material/4595047/systemaattinen_analyysi_tutkimusmetodina.pdf (haettu: 8.12.2019).
- Jones, Thomas M. (1991). Ethical Decision Making by Individuals in Organizations: An Issue-Contingent Model. *The Academy of Management Review*, 16(2), 366–395. DOI 10.2307/258867.
- Jovanovic, B. & Rousseau, P.L. (2005). *General purpose technologies*, National Bureau of Economic Research. Cambridge: Mass.
- Kainiemi, Outi (2019). Vapaudesta aseisiin. *Suomen Kuvalehti*.
<https://suomenkuvalehti.fi/jutut/kotimaa/joukkoampumisten-jalkeen-aseiden-myynti-rajahtaa-yhdysvalloissa-kysynta-saa-asevalmistajat-korottavat-kasiaseiden-hintoja/> (haettu: 8.12.2019).
- Kania, John & Mark R. Kramer (2011). Collective Impact. *Stanford Social Innovation Review* 2011 9(1):36–41.
<http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=55818295&site=ehost-live&scope=site> (haettu 8.12.2019).
- Kansallisen AuroraAI-tekoälyohjelman toimeenpano (s.a.). *Valtiovarainministeriö*.
<https://vm.fi/auroraai> (haettu: 8.12.2019).
- Kant, Immanuel (2004). *Radikaali paha*. Suomentanut Markku Lehtinen, Ari Hirvonen, & Toomas Kotkas. Helsinki: Loki-kirjat.
- Kantele, Tapio (2017). Näin sinua ohjataan Facebookissa ja internetissä. *Yle.fi*.
<https://yle.fi/aihe/artikkeli/2016/12/19/nain-sinua-ohjataan-facebookissa-ja-internetissa> (haettu: 8.12.2019).
- Kauppinen, Antti & Arto Laitinen (2018). Tekoälyn etiikasta – Taustapaperi selontekotyöryhmälle 11.5. *Valtioranainministeriö*.
<https://vm.fi/documents/10623/10841416/Kauppinen-Laitinen-Teko%C3%A4lyn+etiikasta.pdf/d60a235f-9a98-7d22-dac0-1e5b18e3c9ea/Kauppinen-Laitinen-Teko%C3%A4lyn+etiikasta.pdf> (haettu: 8.12.2019).

- Koivisto, Raija, Jaana Leikas, Heidi Auvinen, Ville Vakkuri, Pertti Saariluoma, Jenni Hakkarainen ja Riikka Koulu (2019). Tekoäly viranomaistoiminnassa -eettiset kysymykset ja yhteiskunnallinen hyväksyttävyys. *Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 14/2019*.
<http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161345/14-2019-Tekoaly%20viranomaistoiminnassa.pdf> (haettu: 8.12.2019).
- Korhonen, Suvi (2018). Algoritmit ottavat vallan – hallitsevat ihmisten ajatuksia ja tunteita. *Tivi*. https://www.tivi.fi/Kaikki_uutiset/algoritmit-ottavat-vallan-hallitsevat-ihmisten-ajatuksia-ja-tunteita-6697454 (haettu: 8.12.2019).
- Koutonen, Jouni (2018). Viisi visiota siitä, miten ihminen voidaan tuunata teknologialla. *Yle*.
<https://yle.fi/uutiset/3-10219909> (haettu: 8.12.2019).
- Kokonaisvaltainen ja kestävä systeeminen muutos (s.a.). *Helsingin Yliopisto*, tutkimusryhmät.
<https://www.helsinki.fi/fi/tutkimusryhmat/kokonaisvaltainen-ja-kestava-systeeminen-muutos/tutkimus> (haettu: 8.12.2019).
- Kramer, Mark R. & Mark Pfitzer (2016). The Ecosystem of Shared Value. *Harvard Business Review* 2016/10 80–89. <https://hbr.org/2016/10/the-ecosystem-of-shared-value> (haettu: 8.12.2019).
- Kurzweil, Ray (2001). The Law of Accelerating Returns. *Kurzweilai.net*.
<https://www.kurzweilai.net/the-law-of-accelerating-returns> (haettu: 8.12.2019).
- Laitila, Erkki (2019). *Ihanteeksi vastuullinen tekoäly*. Turku: Painosalama Oy.
- Laitila, Erkki (s.a). Idean Kehittäjä. *Metayliopisto*. <http://metayliopisto.fi/missio/metarehtori/> (haettu: 8.12.2019).
- Leikas, Jaana (2009). Life-Based Design. A holistic approach to designing human-technology interaction [Elämälähtöinen suunnittelu – kokonaisvaltainen lähestymistapa ihmisen ja teknologian vuorovaiku-tussuunnitteluun]. Espoo: *VTT Publications* 726.
- Linturi, Risto & Osmo Kuusi (2018). Suomen Sata Uutta Mahdollisuutta 2018–2037. *Eduskunnan Tulevaisuusvaliokunnan Julkaisu* 1/20.
https://www.eduskunta.fi/FI/tietoeduskunnasta/julkaisut/Documents/tuvj_1%2B2018.pdf (haettu: 8.12.2019).
- Lintilä, Mika (2018). Elinkeinoministeri Mika Lintilän puhe tekoälyn etiikan yrityshaasteen julkistustilaisuudessa. *Työ- ja elinkeinoministeriö*.
https://valtioneuvosto.fi/artikkeli/-/asset_publisher/1410877/elinkeinoministeri-mika-lintilan-puhe-tekoalyn-etiikan-yrityshaasteen-julkistustilaisuudessa (haettu: 8.12.2019).
- Locke, John (1997). *The second treatise of government*. Toimittanut Thomas P. Peardon. New Jersey: Prentice Hall.
- Lu, Yonglong, Nebojsa Nakicenovic, Martin Visbeck & Anne-Sophie Stevance (2015). Five priorities for the UN sustainable development goals. *Nature* 520, (7548) (Apr 23):

- 432–433. <https://search.proquest.com/docview/1675651003?accountid=11365> (haettu: 8.12.2019).
- Microsoft AI Principles (s.a.). *Microsoft.com*. <https://www.microsoft.com/en-us/ai/our-approach-to-ai> (haettu: 8.12.2019).
- Montreal Declaration for Responsible Development of Artificial Intelligence (2018). *ai.Quebec*. https://ai.quebec/wp-content/uploads/sites/2/2018/12/News-release_Launch_Montreal_Declaration_AI-04_12_18.pdf (haettu: 8.12.2019).
- Murphy, R. & David D. Woods (2009). Beyond Asimov: The Three Laws of Responsible Robotics. *Intelligent Systems IEEE* Vol. 24 (4), 14–20. https://www.researchgate.net/publication/224567023_Beyond_Asimov_The_Three_Laws_of_Responsible_Robotics (haettu: 8.12.2019).
- Norvig, Peter & Stewart Russel (2010). *Artificial Intelligence – A Modern Approach*, Third edition. Boston: p., Pearson.
- Novet, Jordan (2018). Facebook forms a special ethics team to prevent bias in its A.I. software. *CNBC*. <https://www.cnbc.com/2018/05/03/facebook-ethics-team-prevents-bias-in-ai-software.html> (haettu: 8.12.2019).
- Nussbaum, Martha C. (2011). *Talouskasvua tärkeämpää*. Suomentanut Timo Soukola. Helsinki: Gaudeamus.
- Ollila, Maija-Riitta (2019). *Tekoälyn etiikkaa*. Helsinki: Otava.
- Oksanen, Markku, Veikko Launis & Seppo Sajama (toim.) (2010). *Etiikan Lukemisto*. Helsinki: Gaudeamus.
- Oksman, R. (2013). Yleiskatsaus projektitutkimukseen. *Projektijohtaminen suomalaisessa tv- ja elokuvatuotannossa*. Luku 2.1, s. 21–29. Helsinki: Aalto University publication series, Doctoral Dissertations 161/2013.
- Packendorff, Johann (2002). The temporary society and its enemies: Projects from an individual perspective. K. Sahlin-Andersson & A. Söderholm (toim.) *Beyond Project Management: New Perspectives on the Temporary-Permanent Dilemma*, 39–58. Malmö: Liber.
- Payment services (PSD2) (s.a.). *Europa.eu*. https://ec.europa.eu/info/law/payment-services-psd-2-directive-eu-2015-2366_en (haettu: 8.12.2019).
- Philström, Sami (2014). Pragmatismi. *Filosofia.fi*. <http://filosofia.fi/node/2409> (Haettu: 8.12.2019).
- Pichai, Sundar (2018). AI at Google: Our Principles. *Blog.google*. <https://www.blog.google/technology/ai/ai-principles/> (haettu: 8.12.2019).
- Pietarinen, Juhani (2015). Etiikka. *Filosofia.fi*. <https://filosofia.fi/node/6985> (haettu: 8.12.2019).

- Platon (2007). *Valtio*. Suomentanut Marja Itkonen-Kaila. Helsinki: Otava.
- Porter, Michael I. & Mark R. Kramer (2011). Creating Shared Value. *Harvard Business Review* 2011/1.
- Public Engagement in Responsible Research and Innovation (s.a.). *Europa.eu – Horizon 2020*. <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/public-engagement-responsible-research-and-innovation> (haettu: 8.12.2019).
- Ray Kurtzweil Biography (2019). *Kurtzweilai.net*. <https://www.kurzweilai.net/ray-kurzweil-biography> (haettu: 8.12.2019).
- Rossi, Venla (2017). Keinoälyfirman perustaja haluaa parantaa syövän ja hidastaa ilmastonmuutosta – ”Tulevina vuosina nähdään läpimurtoja, jotka muuttavat maailmaa enemmän kuin sähkö”. *Helsingin Sanomat*. <https://www.hs.fi/teknologia/art-2000005095803.html> (haettu: 8.12.2019).
- Rousseau, Jean-Jacques (1933). *Émile, eli, Kasvatuksesta*. Suomentanut Jalmari Hahl. Porvoo: WSOY.
- Räisänen, Perttu (2018). Facebook-kohusta kärsinyt Cambridge Analytica hakee konkurssia – yhtiön toiminta loppuu heti. *Kauppalehti*. <https://www.kauppalehti.fi/uutiset/facebook-kohusta-karsinyt-cambridge-analytica-hakee-konkurssia-yhtion-toiminta-loppuu-heti/b0b45d93-9d66-3f5e-8e51-dd1c5335b1f3> (haettu: 8.12.2019).
- Schauer, Michael J. (1995). Estimation of the greenhouse gas externality with uncertainty. *Environmental and Resource Economics* (1995) 5:1 71–82. <https://doi.org/10.1007/BF00691910> (haettu: 8.12.2019).
- Tieteen termipankki (2019). *Nimitys:sosio-tekniinen järjestelmä*. https://tieteentermipankki.fi/wiki/Nimitys:sosio-tekniinen_j%C3%A4rjestelm%C3%A4 (haettu: 8.12.2019).
- Tieteen termipankki (2016) *Filosofia:diskurssietiikka*. <https://tieteentermipankki.fi/wiki/Filosofia:diskurssietiikka> (haettu: 8.12.2019).
- Tieteen termipankki (2016). *Filosofia:instrumentalismi*. <https://tieteentermipankki.fi/wiki/Filosofia:instrumentalismi> (haettu: 8.12.2019).
- Tieteen termipankki (2015). *Filosofia:uusliberalismi*. <https://tieteentermipankki.fi/wiki/Filosofia:uusliberalismi> (haettu: 8.12.2019).
- Waters, Colin N., Jan Zalasiewicz, Colin Summerhayes, Anthony D. Barnosky, Clément Poirier, Agnieszka Gałuszka, Alejandro Cearreta, Matt Edgeworth, Erle C. Ellis, Michael Ellis, Catherine Jeande, Reinhold Leinfelder, J. R. McNeill, Daniel deB. Richter, Will Steffen, James Syvitski, Davor Vidas, Michael Wagemann, Mark Williams, An Zhisheng, Jacques Grinevald, Eric Odada, Naomi Oreskes & Alexander P. Wolfe (2016). The Anthropocene is functionally and stratigraphically

distinct from the Holocene. *Science* 08 Jan 2016: Vol. 351, Issue 6269, aad2622
DOI:10.1126/science.aad2622.

Weeden, Kim A. & David D. Brusky (2013). Inequality and Market Failure. *American Behavioral Scientist* 58/3, 473–491. DOI: 10.1177/0002764213503336.

Wheeler, Gregory (2018). Bounded Rationality. *Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/index.html> (haettu: 8.12.2019).

Wong, Julia C. & Olivia Solon (2018). Google to shut down Google+ after failing to disclose user data leak. *The Guardian*.
<https://www.theguardian.com/technology/2018/oct/08/google-plus-security-breach-wall-street-journal> (haettu: 8.12.2019).

Vainio, Olli-Pekka (2016). Hyve-etiikka. *Logos-ensyklopedia*. <https://filosofia.fi/node/7187>
(haettu: 8.12.2019).

Yleinen tietosuoja-asetus (2019). *Europa.eu*. https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_fi.htm (haettu: 8.12.2019).

LIITTEET

Liite 1. Tekoälystrategiat taulukoituna

| | Suomi | Iso-Britannia | Ranska | Yhdysvallat |
|--|--|---|--|---|
| Tekoälyn kehityksestä vastaa: | Yhteistyössä: <ul style="list-style-type: none"> Kansalaiset Poikkitieteellinen akateeminen tutkimus Yritykset Julkinen hallinto | Yhteistyössä: <ul style="list-style-type: none"> Tekoälyn asiantuntijat STEM* ja dataetiikan aloilta Yritykset Julkinen hallinto | Yhteistyössä: <ul style="list-style-type: none"> Kansalaiset Poikkitieteellinen akateeminen tutkimus Yritykset Julkinen hallinto | Yritykset, joiden toimintaa hallitus rahoittaa. Kehityksen suuntaa ohjaa markkinavoimien näkymätön käsi. |
| Strategian tavoite: | 1.Turvata Suomen talouskehitys. 2.Varmistaa luotettava ja oikeudenmukainen tekoälyn kehitys ja hyödyntäminen. 2.1.Varmistaa kansalaisten valmius tekoälyteknologian leviämiseen. | 1.Varmistaa Ison-Britannian talouskehitys. 2.Tehdä Isosta-Britanniasta houkutteleva valtio tekoälyä kehittäville yrityksille. 3.Mahdollistaa tekoälyyn liittyvien eettisten näkökulmien pohdinta. | 1.Nostaa Ranska tekoälykehityksen ihannevaltioksi (toisin sanoen varmistaa talouskehitys). 2.Varmistaa tekoälyn kehitys ja hyödyntäminen eettisten periaatteiden mukaisesti. 2.1 Mahdollistaa turvallinen innovointi. | 1.Säilyttää Yhdysvaltojen johtopositio tekoälyn kehityksessä ja saavuttaa siitä koituva taloudellinen ja poliittinen hyöty. |
| Tekoälyn kehityksellä ja hyödyntämisellä tavoiteltavat arvot: | Taloudellinen tuottavuus, hyödyn tasainen jakautuminen, vastuullisuus, läpinäkyvyys, yhteiskunnallisen hyväksyttävyyden saavuttaminen. | Tehokkuus, taloudellinen tuottavuus, yhteiskunnallisen hyväksyttävyyden saavuttaminen. | Taloudellinen tuottavuus, hyödyn tasainen jakautuminen, vastuullisuus, läpinäkyvyys, yhteiskunnallisen hyväksyttävyyden saavuttaminen. | Taloudellinen tuottavuus, jonka toivotaan johtavan yhteiskunnallisen hyväksyttävyyden saavuttamiseen. |
| Tekoälykehityksen asettamat uhat: | Työttömyys, lisääntynyt epätasa-arvo, datavääristymät, tekoälyn väärinkäyttö, globaalin poliittisen vallan siirtyminen tekoälyn suuryrityksille, | Tekoälyn väärinkäyttö, Datavääristymät. | Työttömyys, lisääntynyt epätasa-arvo, datavääristymät, tekoälyn väärinkäyttö, vahinkoa tuottava vahva tekoäly. | Osaamisen siirtyminen ulkomaille. Muita uhkia ei voida vielä todeta. |

| | | | | |
|---|---|--|--|--|
| | vallanhimoinen tekoäly. | | | |
| Toimenpiteet tavoitteiden saavuttamiseksi: | Tekoälykoulutuksen implementoiminen kouluasteille, työelämään ja uudelleen kouluttautumiseen. Houkuttimien asettaminen monimuotoisen tekoälyasiantuntijuuden varmistamiseksi. Perustaa parlamentaarinen seurantaryhmä, jonka tavoite on edistää yhteiskunnallista keskustelua tekoälyn kehityksestä ja hyödyntämisestä. Tekoälytutkimuksen rahoittaminen. | Lisätä STEM koulutusta peruskouluihin ja jatko-opintoihin. Rahoittaa STEM + Dataetiikan tutkimusta. Asettaa tekoälyministeri ja –valiokunta pysyväksi osaksi hallitusta. Rahoittaa teknologian kehitystä tukevan infrastruktuurin kehitystä. | Tekoälytutkimuksen rahoittaminen. Houkuttimien asettaminen monimuotoisen tekoälyasiantuntijuuden varmistamiseksi. ”Hiekkalaatikon”** mahdollistaminen innovoimista varten. Tekoälykehitystä ohjeistavan eettisen neuvoston perustaminen. | Juridisten esteiden poistaminen innovoinnin tieltä. Tekoälykehityksen rahoittaminen. Tekoälytutkimusta koordinoivan valiokunnan perustaminen. Koulutusjärjestelmän muuttaminen STEM aineita painottavaksi. |

*STEM tarkoittaa Ison-Britannian hallituksen raportista vastanneiden henkilöiden mukaan tekoälykehityksen näkökulmasta tarpeellisia oppiaineita (science, technology, engineering and maths skills).

**Hiekkalaatikko tarkoittaa tässä yhteydessä rajattua kontekstia, jossa tietyt innovointia haittaavat lainsäädännöt eivät koske tekoälyn kehittäjiä, ja jossa innovointia varten tarvittavat resurssit ovat saatavilla.

Liite 2. Strategioiden tiivistelmät

AI Sector Deal, Iso-Britannia

Hallituksen tulee toimia tekoälykehityksen mahdollistajana, mikä houkuttelee alan osaajia Isoon-Britanniaan ja varmistaa Ison-Britannian talouden vakaan tulevaisuuden.

Mahdollistaminen tarkoittaa mittavia sijoituksia, asiantuntijayhteistyön fasilitointia, osaamisen kehittymisen tukemista ja eettisten näkökulmien pohdintaa.

Tärkeimmät eettiset näkökulmat: Kuinka varmistaa, että käytettävä data ei altistu väärinkäytöksille tai pohjaudu ennakkoluuloille?

Villani Report, Ranska

Hallituksella tulee olla selkeä näkemys, mitä arvoja tekoälyn kehityksellä halutaan saavuttaa ja tukea niiden mukaista toimintaa. Näin tekoäly saavuttaa julkisen hyväksynnän ja Ranskan potentiaali maailman johtavana tekoälyn kehittäjänä voi toteutua.

Tärkeimmät eettiset näkökulmat: Algoritmi neutraaliuuteen ja yksityisen datan käyttöön liittyvät kysymykset ja pohdinta, kuinka kaikki kansalaiset saadaan tekoälystä koituvan hyödyn piiriin.

Tekoälyajan työ, Suomi

Hallituksen tulee tietoisesti edistää arvokkaaksi katsottuja yhteiskunnallisia tavoitteita suunnitellessansa tekoälystrategiaansa ja varmistaa kehityksen läpinäkyvyys, vastuullisuus ja laaja yhteiskunnallinen hyöty. Eettinen tekoälyn kehitys voisi olla suomen seuraava suuri vientituote.

Tärkeimmät eettiset näkökulmat: Kuinka varmistaa, että kansalaiset ovat valmiita tulevaa yhteiskunnallista murrosta varten?

White House summit on Artificial Intelligence for American Industry, Yhdysvallat

Hallituksen tulee varmistaa Yhdysvaltojen säilyminen tekoälykehityksen kärkimaana. Tavoite saavutetaan sijoittamalla rahaa tekoälytutkimukseen ja kansalaisten koulutukseen ja poistamalla tekoälyn kehityksen kannalta haitallisia lainsäädäntöjä.

Tärkeimmät eettiset näkökulmat: Tekoälystä koituvan hyödyn tasainen jakautuminen on tärkeää, jotta tekoälyn kehitys ja hyödyntäminen saavuttavat kansalaisten hyväksynnän. Muutoin hallitus ei lähde taistelemaan mielikuvitushirviöitä vastaan, vaan reagoi, jos ongelmia ilmenee.